DOCUMENT RESUME

ED 373 551 FL 022 351

AUTHOR	Davies, Alan, Ed.; Parkinson, Brian, Ed.
TITLE	Edinburgh Working Papers in Applied Linguistics. Number 5.
INSTITUTION	Edinburgh Univ. (Scotland).
REPORT NO	ISSN-0959-2253
PUB DATE	94
NOTE	134p.; For individual papers, see FL 022 352-359. For earlier volume, see ED 360 833.
AVAILABLE FROM	Edinburgh Working Papers in Applied Linguistics, Institute for Applied Language Studies, 21 Hill Place, Edinburgh EH8 9DP, Scotland, United Kingdom (Individual: surface mail, 5.50 British pounds and air mail 7.50 British pounds; Institutional: surface mail 6.50 British pounds and air mail, 8.50 British pounds).
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Edinburgh Working Papers in Applied Linguistics; n5 1994
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Applied Linguistics; Discourse Analysis; Elementary Secondary Education; Foreign Countries; Higher Education; Language Proficiency; *Language Research; *Language Tests; *Phonology; Pronunciation; *Second Language Instruction; Vocabulary Development
IDENTIFIERS	University of Edinburgh (Scotland)

ABSTRACT

The eight papers in this volume, prepared by staff and students of the Institute for Applied Language Studies of Edinburgh University, address a variety of issues in applied linguistics. The papers include: (1) "A Coding System for Analyzing a Spoken Database" (Joan Cutting); (2) "L2 Perceptual Acquisition: The Effect of Multilingual Linguistic Experience on the Perception of a 'Less Novel' Contrast" (Kayoko Enomoto); (3) "Peer Observation and Post-Lesson Discussion" (Sheena Davies and Brian Parkinson); (4) "Are Score Comparisons Across LP Batteries Justified?: An IELTS-TOEFL Comparability Study" (Ardeshir Geranpayeh); (5) "The University of Edinburgh Test of English at Matriculation: Validation Report" (Tony Lynch); (6) "Some Aspects of 'Foreignness' in the Pronunciation of Upper Intermediate Students of Spanish" (Carmen Santos Maldonado); (7) "Interpreting Metonymy" (Anne Pankhurst); and (8) "Measuring Synonymy as an Intra-Linguistic and Cross-Linguistic Sense Relation" (Matutin Sikogukira). Each paper is followed by a reference list. (MDM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 373 551

EDINBURGH WORKING PAPERS IN APPLIED LINGUISTICS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Brian Parkinson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Number 5
1994

BEST COPY AVAILABLE

FL022351

EDINBURGH WORKING PAPERS IN APPLIED LINGUISTICS

Number 5

1994

(ISSN 0959-2253)

Published jointly by the Department of Applied Linguistics
and the Institute for Applied Language Studies
University of Edinburgh
Scotland

Editors Alan Davies
Brian Parkinson

Subscription information

Edinburgh Working Papers in Applied Linguistics (EWPAL) is published in the third term of the University of Edinburgh year (April-June). Details of subscription rates and journal exchange arrangements are at the end of this issue.

Copyright

Copyright rests with the authors of the papers in this volume, who are free to submit their work for publication elsewhere, either as published here or in a revised version.

Contents

A CODING SYSTEM FOR ANALYSING A SPOKEN DATABASE
Joan Cutting 1

L2 PERCEPTUAL ACQUISITION: THE EFFECT OF MULTILINGUAL LINGUISTIC EXPERIENCE ON THE PERCEPTION OF A "LESS NOVEL" CONTRAST
Kayoko Enomoto 15

PEER OBSERVATION AND POST-LESSON DISCUSSION
Sheena Davies and Brian Parkinson 30

ARE SCORE COMPARISONS ACROSS LP BATTERIES JUSTIFIED?: AN IELTS-TOEFL COMPARABILITY STUDY
Ardeshir Geranpayeh 50

THE UNIVERSITY OF EDINBURGH TEST OF ENGLISH AT MATRICULATION: VALIDATION REPORT
Tony Lynch 66

SOME ASPECTS OF 'FOREIGNNESS' IN THE PRONUNCIATION OF UPPER INTERMEDIATE ST[illegible]TS OF SPANISH
Carmen Santos Maldonado 78

INTERPRETING METONYMY
Anne Pankhurst 98

MEASURING SYNONYMY AS AN INTRA-LINGUISTIC AND CROSS-LINGUISTIC SENSE RELATION
Matutin Sikogukira 109

Preface

The eight papers in EWPAL 5 provide a good indication of the variety of research in applied linguistics currently being undertaken in the University of Edinburgh. The issues addressed include: language proficiency testing (two papers), stylistics, vocabulary, discourse, classroom observation, pronunciation and phonological acquisition (one paper each). Six of the papers have been written by current research students, two by staff of IALS. Since, however, several of these research students are also involved in IALS teaching and the three IALS authors all teach in the Department (and two are former students), it is happily unnecessary to make so categorical an institutional distinction. To put this another way, what EWPAL 5 represents is exactly what the publication was set up to achieve, a bringing together of the Department and the Institute by sharing research.

I would like to thank the following for reviewing the submissions to EWPAL 5: Cathy Benson, Joan Cutting, Esther Daborn, Ardeshir Geranpayeh, Martin Gill, Eric Glendinning, Phillip Goertzen, Tony Lynch, Joan Maclean, Keith Mitchell, Elni Rigas, Dan Robertson, Liam Rodger, Sonia Shiri, Antonella Sorace.

Brian Parkinson and I have shared editorial responsibility for this issue. Since I was away from Edinburgh at the latter stage of the production of EWPAL 4, Brian took on most responsibility for that issue and I would here like to acknowledge my debt to him for taking on so much last year. I must also express the thanks of the Department of Applied Linguistics to the Institute for Applied Language Studies for taking on once again the tasks of desk top publishing and distribution. While the Department does make a contribution to the costs of printing, the Institute continues to pay a larger share of the total cost. That needs to be stated.

Thanks are due to Elaine Bell of IALS who has once again speedily and efficiently produced the camera-ready text from contributors' 'final' versions, and to Ray Harris and colleagues of the Reprographics Department of the University for producing the published version of EWPAL 5.

Alan Davies

May 1994

A CODING SYSTEM FOR ANALYSING A SPOKEN TEXT DATABASE

Joan Cutting (DAL)

Abstract

This paper describes a coding system devised to analyse conversations of 1991-92 Applied Linguistics MSc students at Edinburgh University. It gives details of the lexical and grammatical tags that are applied to the text itself and outlines the code for the analysis of each discourse unit's assumed knowledge area, macro-function, speech act, and move. The problems encountered in the implementation of this system are discussed and explained, and solutions are offered.

1. Introduction

There have been several studies, in the fields of sociolinguistics and psycholinguistics, that compare the language of strangers and that of friends (Tannen 1989; Duck 1991). However, it would appear that there is no longitudinal linguistic study that follows through the interactions of speakers from the moment that they first meet until they become friends or associates, in order to discover exactly how their language develops over time.

I am making a developmental study of casual conversations of MSc students from the 1991-92 Edinburgh University Applied Linguistics course, aiming to find exactly how their language changes as they form a discourse community. The analysis is hoped to explain how the in-group's code becomes increasingly restricted (Bernstein 1971) and thus inaccessible to an outsider to this MSc group, and to provide a model for predicting language changes during discourse community formation. It is not an *a priori* model: the categories have been devised as a result of examining the dialogues with an ethnomethodological eye.

In order that the pragmatic analysis of the implicit language that evolves in this closed network academic group (Levy 1979; Kreckel 1981) can be both qualitative and quantitative, I have devised a detailed coding system. The system contains certain lexical and grammatical features that depend on the context of the MSc course for their full meaning: special terms and names, general words (Halliday and Hasan 1976) and exophoric reference, substitution and ellipsis. I call these features implicit 'contextualisation cues', to use Gumperz's term for linguistic features that contribute to the 'signalling of contextual presuppositions' (1982:71). I hypothesise that over time, as the group becomes closer knit, united by the common experience of the MSc course, there will be an increase in implicit contextualisation cues.

The coding system also contains functional categories, since the linguistic features must be seen in the context of when, how and why they are used, so that a statement may be made about group dynamics. Thus I analyse each discourse unit in terms of perlocutionary function, implicature and speech act, as well as the knowledge area assumed.

This paper describes each category of my coding system. It discusses the difficulties involved in making the system work and adapting the categories to meet the needs of natural language.

2. The Overall Methodology

2.1 Hypotheses

I hypothesise that as shared knowledge grows, the textual density of implicit contextualisation cues increases and that the language of in-group members has more cues than that of strangers. I hypothesise that course-related topics will be more impenetrable than non-course-related topics to an outsider to the course; that course-related topics will become more frequent than non-course-related topics; and that this will cause the conversations to have larger impenetrable sections because of both the assumed knowledge area and the density of implicit contextualisation cues.

My hypothesis about the pattern of development of the implicit contextualisation cues is that after the beginning of the MSc course, there will be a peak of special terms (eg: 'X-bar'), proper names (eg: 'Chomsky'), demonstrative and comparative reference (eg: 'this', 'more'), combined with a drop in explicit endophoric noun phrases with post-head dependents (eg: 'that we did', 'in class'). As the course progresses, special terms and names will level off and there will be an increase in third person personals (eg: 'she'), indefinite pronouns (eg: 'anybody'), exophoric substitution (eg: 'the one') and ellipsis, and superordinates (eg: 'book'), and general nouns and verbs (eg: 'thingy','do'"). This overall trend will be affected by events such as portfolio dates and project deadlines, which lead to conversations laden with technical terms and proper names.

I predict that certain informal aspects of the language will increase as members become more familiar with each other: sociocentric fillers such as 'you know' and 'I mean', slang and expletives, and sentences with no initial subject or auxiliary and unfinished sentences.

I take into account two secondary but essential factors: cohesion and function. A consideration of cohesion should reveal that as the language becomes more exophoric, general and bald, as explicitness is no longer necessary, the risk of breakdowns and requests for clarification increase, especially in course-related topics, because speakers wrongly assume that all the relevant common knowledge is in their hearer's mind.

The analysis of the function of utterances containing cues should show that the use of implicit contextualisation cues is a generally expected unmarked means of claiming in-group membership (Levinson 1978; Tannen 1989). I hypothesise that,

in both transactional and interactional social exchanges, there will be an increase in the students' manipulation of shared knowledge as they flout quality maxims. I predict, too, an increase in speech acts that are demonstrations of 'in-the-same-boatness', such as expressions of dissatisfaction with self or situation, or positive evaluation of the interlocutor.

2.2 Method of data collection

I openly made 15 tape-recordings (a total of 4 hours 40 minutes) of MSc student conversations in the common room of the Applied Linguistics department from 4 October 1991 until 12 May 1992. I recorded once a week over three periods of time: the first half of the first, second and third term. The conversations (29 in total) were spontaneous and unguided, and I kept at a distance at the moment of recording so as not to be included. Six native speakers of English who had options in common and tended to sit together in the common room consistently were, four weeks into the course, selected for analysis, on the assumption that they were representative of all native speakers of English on the course. Once I had transcribed the recordings (26,000 words), I disregarded dialogues or long sections of dialogues which did not contain at least two of the six chosen speakers.

2.3 Method of data analysis

I analysed all discourse units of the text (3,500) in terms of knowledge area and function, in six fields. I also tagged the text (Field seven) lexically and grammatically.

Figure One: Fields and tags of the coding system

Functional Fields:		
1)	Dialogue and discourse unit numbers	
2)	Knowledge area	
3)	Macro-function	
4)	Move and topic	
5)	Speech act	
6)	Speakers	
Text Field and linguistic tags		
7)	Lexical :	verbs
		nouns
	Grammatical:	articles
		pronouns and adjectives
		substitution and ellipsis
		dependents

The six functional fields were dialogue number and discourse unit number, knowledge area, move and topic, speech act, and speakers. Of the text tags, the lexical ones were mainly verbs and nouns, tagged with numbers, and the grammatical ones were articles, pronouns and adjectives, substitution and ellipsis, and dependents tagged with letters/symbols.

Let us now examine the individual categorisation of these functional labels and text tags. All examples quoted in this paper are taken from my database. The numbers

beside examples quoted here are from Field one: dialogue and discourse unit. The first two digits indicate the dialogue number; the other three indicate discourse unit number. The letters are from Field six: the speakers. AM, BM, CM, DM were the men selected for study in this research, and AF, BF the women.

3. The Individual Categorisation

3.1 The discourse unit labels

3.1.1 Field one: dialogue and discourse unit.

My criteria for deciding what constituted a discourse unit were as follows: that a unit can end where a grammatical sentence ends, regardless of whether this is the end of a turn or move; that a unit always ends where there is falling intonation combined with a pause longer than 0.5 seconds, even though it happens in the middle of a sentence; and that each discourse unit constitutes a speech act. Any utterance occurring simultaneously, unless it is unintelligible, is considered a unit.

3.1.2 Field two: knowledge area

I divided assumed common knowledge into four areas, in order to distinguish non-course-related topics (areas 1, 2 and 3) from course-related topics (area 4). I added the symbol P if I felt there was shared, privileged or interpersonal knowledge in any of the four knowledge areas.

Figure Two: Knowledge areas

1	The world, Edinburgh
2	Linguistics, language teaching and learning
3	Edinburgh University, Department of Applied Linguistics, Institute for Applied Language Studies,
4	The 91-92 MSc course in Applied Linguistics of Edinburgh University
(P	Privileged or shared knowledge in any of the above four areas)

When I was assigning discourse units to the four knowledge areas, I began to feel that I only needed two categories: course-related and non-course-related. However, since some topics were more course-related than others and some were course-related topics but also related to the 'real world' outside, I maintained the original four areas. Because it was hard to say where one knowledge area ended and another began, at the discourse unit level, a knowledge area was established for the duration of a whole topic, wherever possible.

3.1.3 Field three: macro-function

Discourse units were next labelled T for the transactional function or S for the social, the interactional. I did not predict an increase or decrease in either over time; I wanted to examine the grammatical and lexical tags within the two

functions. Again, the overall function was established for a whole topic, rather than on an individual discourse unit level.

Figure Three: Macro-functions

T	Transactional, instrumental
S	Interactional, social
(F	Flouting of quality and clarity maxims in either of the above two)

T was the purely transactional functional category, the instrumental such as,

e.g. 01054 BM 'Could you get me a tuna and sweet corn one please?'
01055 AM 'Me as well.'

or learning specifically how to do something, transmitting information needed for an immediate task, such as

e.g. 06011 CM 'The best thing to do is use a che- master card or something.'
06012 NF 'I don't have a master card.'

All other units were tagged with S for the social expressive exchange, the interactional category, and the phatic. In many cases, the speakers were testing the normality of their situation or feelings.

e.g. 15020 CM 'But I suppose you're moderately efficient. (1)'
15021 CM 'But I'm not near an outline.'
15022 DM 'Oh God no I'm on my reading.'

In others, they were reassuring themselves that they had the same information and checking each other's attitude to it.

e.g. 08063 CM 'Five's a bad mark right?'
08064 AM 'I think we get five we're fine.'
08065 CM 'It's the other mark I want.'

There is a cline from the instrumental to the purely phatic. Even the purely phatic could have a tinge of information exchange. Sometimes an interaction that, on the surface, was an information exchange could be fundamentally a social exchange.

In the same field, I added a label F to indicate that the speaker was playing with truth and/or the language, flouting cooperative maxims by exaggeration, irony, banter; playing with clichés in mock seriousness, or mixing registers to amuse etc. I wanted to test whether humorous exchanges were a marker of in-groupness that increased over time.

e.g. 04080 BM 'They're sort of we give change!'
04081 DM 'If you can get on the bus you get change then.'
04082 BF 'Ah right // that's good.'
04083 DM '// Yeah.'
04084 AF 'If you don't mind breaking your leg as you try and get on.'

3.1.4 Field four: topics and moves

I was especially interested in examining topic shifts to see how often implicit contextualisation cues occurred in the first discourse unit of a new topic. I studied patterns of moves, to see whether this revealed something about group dynamics.

Figure Four: Topics and moves

T	Initiate
R	Respond
1	1st. discourse unit of/within a turn, on a new topic
2	Subsequent discourse units within a turn, on an established topic
3	1st. discourse unit of a turn, on an established topic

'Topic' I defined as part of an interaction that can be given a title in the form of an indirect question that covers all speakers' contributions to it. Thus an example of this discourse topic would be 'Why X is not going to do any reading' or 'What Y did on his mountaineering weekend.' To analyse topics, I tagged the first discourse unit of a turn or within a turn on a new topic '1', I tagged subsequent discourse units within a turn on same topic, even if still answering the same question '2', and the first discourse unit of a turn on existing topic '3'. When the topic drift was so gentle that it was almost imperceptible, I placed topic boundaries as close to the centre of the shift as possible.

I classed the moves T for Initiate and R for Respond. R was either the second half of an adjacency pair or simply a minimal response to the preceding discourse, such as a backchannel or prompt to continue. T was everything else. T could occur in T1, T2 and T3. R tended to occur mainly in R3, as in the first discourse unit of an answer or adjacency pair response. R2 occurred rarely because subsequent discourse units of an answer were not normally specifically required by the question.

e.g. -11105 NM 'You went home?'
11106 AM 'I went home round about half past seven no.'
11107 AM 'I didn't go.'

11106 was, strictly speaking, all that was needed to answer NM's question. 11107 was T2 rather than R2. Finally, R1 was the change of focus to the metalinguistic, the request for clarification, or the answer to a question from a previous topic, ignoring an interspersed topic.

3.1.5 Field five: speech act

The task of classing discourse units in speech acts was particularly daunting because speech acts are so elusive and one unit often fits into more than one speech act category. When a unit did not fit neatly into an act category, I created a new category, or expanded an existing category, and noted this realisation of the act in order to guarantee consistency.

I initially detected 40 speech acts, but having failed to make them water-tight and mutually exclusive, I opted for grouping them into nine macro speech acts with no subdivisions. Discourse units were analysed in terms of attitude (neutral, positive and negative) and then in terms of person or object discussed (self, interlocutor or communication, and third party or situation). This solution obviously did not produce a delicate tool but I felt that it would be more manageable for the quantitative analysis stage.

Figure Five: Speech acts

attitude towards:	Neutral	Positive	Negative
Self	11 I do/am	12 I do/am good	13 I do/am bad
Interlocutor/ communication	21 You do/are Here we are	22 You're good I'm with you	23 You're bad I'm not with you
Third party/ situation	31 It/she does/is	32 It/she does/is good	33 It/she does/is bad

This paper permits no more than a general view of each macro speech act. As far as Attitude towards Self was concerned, within act 11, I included inform, explain, answer and express own intentions and desires; within act 12, express satisfaction with self, and reassure and console self; within act 13, express dissatisfaction with self, minimise praise of self, excuse self and apologise.

Similarly, for Attitude towards Interlocutor or Communication, there was in act 21, greet, fill phatically with afterthoughts, backchannelling, laughter etc., request; in act 22 evaluate interlocutor positively, console, encourage, sympathise, agree, predict drift, advise and suggest, offer promise goods or action; and in act 23, challenge factual content/truth value, decline, and deny.

Finally, in Attitude towards Third Party or Situation, I included in act 31, inform, explain, ask/answer; in act 32, express satisfaction with third party or situation; and in act 33, express dissatisfaction with them, express fear and apprehension. And thus I hoped to cover everything that MSc students do with language in the common room.

3.2 Field seven: the tagged text

3.2.1 Lexical tags

Although the main emphasis of my study is the noun phrase, I included an analysis of all verbs, because I wanted to show that over time there would be an increase in general verbs 'do' and 'make', in terms of a percentage out of all verbs.

e.g. -10069 CM 'I've done all the people'

I had one category for the course-related general verb and one for the non-course-related.

The other category of verb that I tagged separately from the 'all verbs' category was the filler: both sociocentric sequences (Stubbs 1983) such as 'I mean', 'You see', and omissible hedging verbs with or without sentential objects (Brown and Miller 1980) such as 'I suppose.'

e.g. -17020 AF 'It is a lot you know three thousand words'.

Figure Six: Lexical tags - verbs

10	All Verbs (excluding be, filler, general verbs)	eg: 'bought'
11	Be	eg: 'is'
12	Filler	eg: 'I mean'
13	Non-course general verb	eg: 'do'
14	Course general verb	eg: 'do'

In my categorisation of nouns, I tagged common nouns, proper nouns and general nouns, indicating whether the referent was course-related or non-course-related.

I tagged expletives and slang, being interested in these as markers of intimacy that might increase in number with repeated interaction. I included certain adjectives:

e.g. -08020 AM 'Bloody tosh, isn't it?'
08021 CM 'Well it's- it's a bit abstract. ((1))'

within the group of expletives, and phrasal expressions:

e.g. -10029 DM 'So what on earth is it going to be about?'

within the slang group.

Figure Seven: Lexical tags - nouns

''	Expletives	eg: 'hell'
''	Slang	eg: 'scivers'
20-22	Non-course nouns	
20	Common non-course nouns	eg: 'budgie'
21	Proper non-course names	eg: 'Japan'
22	General non-course nouns	eg: 'thing'
30-50	Course nouns - special	
30-32	Common special course nouns	
30	technical	eg: 'X-bar'
31	unique	eg: 'portfolio'
32	superordinate	eg: '[syntax] book'
40-41	Proper special course names	
40	actual use	eg: 'Chomsky'
41	metonymical use	eg: 'Chomsky [study]'
50	General special course nouns	eg: 'thing'
60	Course nouns - course-by-context	eg: 'work [for the exam]'

Categorising nouns with non-course-related referents was quite simple. Words such as 'budgie' and 'bus-stop' were classed as common nouns, 'Bush' and 'Manchester' as proper nouns, and 'thing' and 'people' as general nouns.

Classing nouns with course-related referents was more complex. These were divided into two main groups: special course nouns and course-by-context nouns. The special course nouns, intrinsically course-related, were classed under the headings common, proper and general nouns.

Under the first special course nouns heading, that of common nouns, I made three sub-divisions - technical, unique and superordinate. I tagged nouns technical if they were intrinsically specialised terms independent of the context of the course, linguistics and language teaching theory terminology such as 'X-bar', 'diglossia' and 'lexical syllabus'. I tagged nouns unique common special course nouns if they were single course components such as 'portfolio', 'core project', and 'the examiner', used in department organisation, the course handbook etc. Superordinate common special course nouns constituted a class of many possible course components. Superordinates were count nouns such as '[syntax] book', '[SLA] class' and '[tutorial] task' whose precise meaning was not clear since they were the second noun of a two-word phrasal expression (Huddleston 1988:103) whose first word (usually a nominal pre-head modifier) was omitted. In this

e.g. 20020 CM 'SLA class next week and the paper's due in next Friday'

the paper was the core project, the first project paper of the course.

Under the proper special course noun heading there were two sub-divisions: one for proper nouns with their actual intended use, such as 'Chomsky', 'ELT' and 'Structuralism', and the other for their metonymical use, in which the noun referred elliptically to something other than the course/linguist named. As with the superordinate common special nouns, these nouns could be seen as one word of a two-word phrasal expression, except that in the case of the metonymical use, it is the second word that is omitted. In

e.g. 08031 AM 'Though though I haven't I haven't done any Chomsky.'

'Chomsky' meant Chomsky study or revision for the forthcoming examination.

The general special course noun category was structurally the same as the general non-course noun. To take an example:

e.g. 10069 CM 'No I've done all the people.'

Here, 'people' meant theorists or linguists.

The second main group of course nouns I called course-by-context nouns. These were not intrinsically course-related but became course-related by their use in course-related topics. The referents of 'discussion', 'work' and 'this week' were course-related, for instance, when the 'discussion' was one that the students had had in a tutorial, 'work' meant work for the examination, or 'this week' meant this

BEST COPY AVAILABLE 14

week in classes. Here what seemed to be missing was a post-head dependent such as a relative clause, a prepositional phrase or a content clause.

3.2.2 Grammatical tags

Grammatical tags were put on articles, pronouns and adjectives, substitution and ellipsis, and dependents.

Figure Eight: Grammatical tags - articles, pronouns, adjectives, substitution and ellipsis

Articles		
T/A	indefinite article / definite article	
Pronouns and adjectives		
P	Indefinite pronoun	eg 'anybody'
E	Existential pronoun	eg 'she'
S	Possessive adjective	eg 'their'
D	Demonstrative pronoun / adjective	eg 'this'
C	Comparative pronoun / adjective	eg 'more'

I tagged all articles, predicting that there would be an increase in the partially explicit exophoric use definite noun phrase, especially with course-related nouns (including general course-nouns), as in this example:

e.g. -16032 AF 'I think- I find a real loss actually of not having read the (0.5) Fay Cutler article which seems to be underpinning this ='

I tagged the indefinite pronouns 'everything', 'nobody' etc., feeling that they have an exophoric quality. They constituted a 'vague' feature of in-group language. In the following example:

e.g. -25020 NF 'Is there anybody else applying?'
the full extent of 'anybody else' could only be understood by group members.

The tagging of pronouns, adjectives, substitution and ellipsis proved to be a complex task. There were different degrees of endophora and exophora. I considered pronouns and adjectives endophoric if the referent was in the text, whether it was an item or a proposition, even though the reference would have been unclear or ambiguous to our hypothetical outsider because of the existence of a number of possible textual referents. I decided to tag as semi-implicit endophoric, the adjectives that modified general nouns and the pronouns and substitutes that had endophoric cohesion to general nouns, exophoric pronouns or substitutes, or indefinite pronouns. On the borderline between endophora and exophora seemed to be associative anaphora in which the referent could be understood from the schemata established by other referents in the text.

The rest of the pronouns, adjectives, substitutes and ellipses were exophoric or non-anaphoric. Here, the degrees of implicitness were more noticeable. Again I differentiated between adjectives that modified common nouns explicitly, and those that could be called semi-implicit exophoric because they modified general nouns

or exophoric substitutes. Most demonstratives, for example, are exophoric, but demonstrative adjectives seemed more implicitly exophoric when they modified the impenetrable, intertextual general nouns or exophoric substitutes.

e.g. -15045 DM 'So I typed that thing up again after you'd gone.'

The pronouns, substitutes and ellipses whose referents were irrecoverably right outside the text were classed as totally implicit exophoric.

Tagging pronoun and adjective categories revealed interesting particularities of each. Demonstratives seemed more obviously intertextual than other forms of reference. In the following example, the 'pizza' outing had evidently been discussed on a former occasion, yet it was exophoric as far as the text was concerned.

e.g. 11089 AM 'Did you go to this pizza on Friday?'

Some of the characteristics of demonstratives were typical of spoken language. I tagged the demonstrative pronoun in the expression 'like that' in an unfinished sentence not as a demonstrative but as a substitute expression, or 'vague expression' (Channell 1985). Familiarity with the presuppositional pool is needed to complete the sentence.

e.g. -04099 BF 'And King's Building's got a bar and stuff like that.'

Narrative sections featured the informal demonstrative adjective 'this', which brings the story closer to the hearers.

e.g. -02084 MM 'So so he keeps he's got this really incredible part.'

Comparative reference has three points of reference: the referent mentioned, the referent that it is compared to, and the quality that unites or divides them. I was tempted to tag some of the cases in my data exophoric, because of the exophoric comparing quality.

e.g. -15076 DM 'The actual thing's exactly the same.'

I opted not to call these cases exophoric, however, since the 'exophoric-ness' of a quality was too flimsy to pin down.

Two final points need to be made in connection with tagging pronouns and adjectives in spontaneous recordings of groups of three or more speakers. The first is that since few recordings contained beginnings of conversations, most conversations were captured mid-topic. Thus there were reference items that looked as if they might have been endophoric, with the referent probably in the preceding, unrecorded, stretch of conversation. Faced with this uncertainty, I tagged such cases using guesswork based on the text.

The second point is that some speakers missed the textual referents because they joined conversations already under way. Thus, although I might tag one item of reference as endophoric, it could seem exophoric to some hearers. Witness:

e.g. -11064 BF 'And they were doing a profile of him.'
11065 BF 'And I was sort of =.'
11066 NM 'Who was it?'
11067 BF 'He was a guy who wrote this dirty book.'
11068 BF 'Melvyn Bragg?'

Here, BF had to supply the missing restrictive post-head dependent and proper name, to repair the breakdown.

Tagging substitution and ellipsis, I found that exophoric substitution was predictably rare, apart from the vague substitute expression at the end of the unfinished sentence, and that ellipsis was more interesting to tag as there were more instances of implicit use. Here is an example of semi-implicit endophoric: the 'them' in line 25014 is exophoric.

e.g. -25014 CM 'So what are em how many of them are there do you know?'
25015 BF 'I don't know. (0.5)'
25016 BF 'Don't know.'
25017 BF 'Em.'
25018 BF 'I think (1) three or two ... I'm not sure.'

I tagged ellipsis of initial subject or auxiliary, interested in it as an example of speakers' inexplicitness about the "given" part of the information structure. This informal style could be a marker of intimacy.

e.g. -24040 DM 'Something wrong isn't there somewhere?'

The ultimate ellipsis is the unfinished sentence: the speaker assumes that the hearer can supply the missing information.

e.g. -26131 BM '// Have you got this on the the drive at the...?'
26132 CM 'No // no you can't do that.'

I call an unfinished sentence one that is deliberately left 'in the air', and not one that is unfinished because another speaker interrupts.

The last grammatical elements, dependents, were fairly straightforward to tag. I predicted that as assumed common knowledge increased, there would be less need to identify the nouns with a description, whether in the form of a pre-head dependent (determiner or modifier) or a post-head dependent (complement, modifier and peripheral). Post-head dependents will abound in the first half of the course:

eg. 16088 AF 'I mean it all to do depends on the point of view of the research which we're not really // adequate to judge.'

whereas, by the end, the hearer will probably have to supply them:

eg. 27128 DM 'If you're answering questions anyway so (3) shouldn't be a problem. ((2))'

BEST COPY AVAILABLE

Figure Nine: Grammatical tags - dependents

Pre-head	determiner, modifier
Post-head	complement, modifier, peripheral

4. Conclusion

4.1 Summary and comments

This paper has examined a coding system for analysing the recorded casual conversations collected in the 1991-92 Applied Linguistics MSc common room. It has explained the system of six functional fields to analyse the speech acts and knowledge areas of each discourse unit, and one actual text field with lexical and grammatical tags. The problems of each have been detailed and the solutions described.

Although I first constructed the coding system using the information that I had gained from an impressionistic global survey of the data, once I was actually using the code to analyse the dialogues, I had to adjust it constantly in response to the dictates of the data. On finishing the coding, I had to adapt and re-shape the system again to make all the categories fit together into a coherent model. Finally, I have confirmed the workability of the code, using intercoder reliability tests.

4.2 The next stage in the research

Using the Excel database, the next stage is to count the implicit contextualisation cues and other linguistic items, as well as the discourse units in the function fields of knowledge area, macro-function, topic shifts and moves, and speech acts, in each dialogue; and to find the average percentage of these for each of the three recording periods of the course, and thus observe whether there are any obvious changes over time: whether knowledge area four topics do increase; whether the number of cues does increase; which cues increase at which point; whether language becomes more informal, and so on. The second stage is to carry out a qualitative analysis of chunks of dialogue or even whole dialogues to find characteristics of the ways that cues relate to each other within a text.

The third stage is to put all the information together, to observe linguistic items in conjunction with the various functional fields and discover whether their density varies according to the knowledge area, the macro-function and the speech act. This would test the hypotheses that course-related topics have more implicit, impenetrable language; and that the use of cues is related to the claiming of in-group membership. Impenetrability itself will be tested through questionnaires given to speakers from each of the four knowledge areas.

The subsequent statistical analysis should help me to select the most significant implicit contextualisation cues to include in the final model which should suit any body of data taken from conversations between in-group members of an academic discourse community. I hope eventually to make the model generalisable to English-speaking academic discourse communities in Britain, and possibly non-

academic discourse communities to explain why their conversations can sometimes be impenetrable to outsiders.

References

Bernstein B. 1971. Class, Codes and Control. Vol 1. London: Routledge and Kegan Paul.

Brown E. and J. Miller. 1980. Syntax. A Linguistic Introduction to Sentence Structure. London: Hutcheson University Library.

Brown P. and S. Levinson. 1978. Politeness. Cambridge: CUP.

Channell J. 1985. 'Vagueness as a conversational strategy' in Nottingham Linguistic Circular. 14: 3-24.

Duck S. 1991. Friends For Life. Hertfordshire: Harvester Wheatsheaf.

Gumperz J. 1982. Discourse Strategies. Cambridge: CUP.

Halliday M. and R. Hasan. 1976. Cohesion In English. London: Longman.

Huddleston R. 1988. English Grammar. An Outline. Cambridge: CUP.

Kreckel M. 1981. Communicative Acts and Shared Knowledge in Natural Discourse. London: Academic Press.

Levinson S. 1983. Pragmatics. Cambridge: CUP.

Levy D. 1979. 'Communicative goals and strategies: between discourse and syntax' in T. Givon (ed.) 1979. Syntax and Semantics 12. 183-210.

McCarthy M. 1991. Discourse Analysis for Language Teachers. Cambridge: CUP.

Stubbs M. 1983. Discourse Analysis. Oxford: Basil Blackwell.

Swales J. 1990. Genre Analysis. Cambridge: CUP.

Tannen D. 1989. Talking Voices. Cambridge: CUP.

BEST COPY AVAILABLE

L2 PERCEPTUAL ACQUISITION: THE EFFECT OF MULTILINGUAL LINGUISTIC EXPERIENCE ON THE PERCEPTION OF A "LESS NOVEL" CONTRAST

Kayoko Enomoto (DAL)

Abstract

This paper reports upon a small-scale L2 perception study on the acquisition of stop durational contrasts in Japanese. The study was designed to investigate the effect of specific vs. non-specific, broadened multilingual linguistic/perceptual experience on the learner's perceptual ability to discriminate between single and geminate stops in Japanese. To explore this, an identification test and an AXB discrimination test were conducted with two different types of multilingual adult learners of Japanese and their English-speaking monolingual counterparts. Amongst the multilingual subjects, some spoke languages which utilise segmental durations phonemically, whilst some spoke languages which do not. Whilst overall data from both tests indicated superior perceptual performance by the group of multilingual learners, there was no significant difference in perceptual performance between the two types of multilingual subjects. Thus, the results support the hypothesis that not only specific but also non-specific, broadened linguistic experience can result in increased perceptual performance. In addition, the overall pattern of performance by monolingual subjects was superior to the patterns reported in other studies, as well as indicating their existing perceptual sensitivities towards phonetic differences in the acoustic cue within the same category. These results and their theoretical implications are discussed.

1. Infant speech perception

Studies of speech perception by infants developed from research focussing on the relation of the speech signal to phonemes of language, most commonly concerned with stop consonants varying in voice onset time (VOT)[1]. Most work on infant speech perception has been designed to test L1 perception rather than cross-language perception. The classic experiment by Eimas et al. (1971) described the ability of infants to perceive speech signals as deriving from genetic predisposition. This nativistic view was subsequently challenged by the view that phonemic categories are determined by early linguistic experience in a particular language environment. In support of this proposition, Aslin et al. (1981) demonstrated that infants from an English-speaking environment can reliably discriminate an irrelevant VOT contrast, occurring within the same English-adult phonemic categories.

Following this pattern of findings, Eimas (1985), in his subsequent study, claims that human beings are endowed with innate perceptual mechanisms which facilitate the acquisition of a language.

> In the studies of speech perception by infants we have found these young subjects are richly endowed with innate perceptual mechanisms, well adapted to the characteristics of human language, that prepare them for the linguistic world they will encounter.
>
> (Eimas 1985:34)

This proposition derives from the view, advanced by Chomsky (1981), that innate knowledge and capacities underlie the use of language, whereby infants are born with innate perceptual capacities/sensitivities which enable them to discriminate between the universal set of phonetic distinctions, according to universal phonetic boundaries; if these perceptual mechanisms do represent an innate biological endowment, they should be universal. The same perceptual patterns should occur in infants of every linguistic background. For instance, it may be hypothesised that Japanese infants have an innate sensitivity towards the English /l/-/r/ distinction, whilst many studies (e.g. Goto 1971; Miyawaki et al. 1975; Mochizuki 1981) report that adult native Japanese speakers fail to perceive the contrast.

Thus, it may be argued that there is a decline or loss in initial perceptual capabilities after being exposed to a particular L1 which presents distinctions only in certain contrasts. To support this proposition, cross-language speech perception studies, focussing on the relation between infant and adult perceptual categories, have addressed the question of adult-infant differences by directly comparing their ability to discriminate non-native contrasts. For instance, Trehub (1976) showed that, whilst English-speaking adults achieved perfect accuracy with English contrasts, they were constantly confused with French and Czech contrasts. On the other hand, infants could differentiate both contrasts which they are not likely ever to have heard before. Furthermore, Werker et al. (1981) found such a developmental decline between infancy and adulthood.

2. The effect of linguistic experience

The findings from cross-language speech perception studies suggest the need for further inquiry with regard to the decline or loss of initial perceptual capabilities in adult speech perception.

2.1 The effect of training

Laboratory training studies have reported that, after intensive training, Japanese learners of English showed some success in improving the perception of /l/-/r/ (for discussion of the methods and tasks used in such training, see Gillette 1980; Strange and Dittmann 1984; Logan et al. 1991). These studies on /l/-/r/ suggest that the training process requires intensive instruction and considerable time and effort, at least for some types of phonetic contrasts.

> It was concluded that modification of perception of some phonetic contrasts in adulthood is slow and effortful, but that improved laboratory training tasks may be useful in establishing categorical perception of these contrasts.
> (Strange and Dittmann 1984: 131)

Likewise, MacKain et al. (1981), in a cross-sectional study, also found that intensive conversational instruction with native speakers correlated with improved perception of /l/-/r/ by Japanese learners of English, although the perception of the experienced Japanese group was not yet native-like. On the other hand, other training studies employing synthetic VOT continua in general showed rapid improvement with relatively little training (e.g. Pisoni et al. 1982; McClaskey et al. 1983).

These results, taken as a whole, may be interpreted as indicating that innate perceptual mechanisms which formerly enabled infants to make universal distinctions are still available to adults and operate in adulthood after a long period of not being used. This confounds the hypothesis that early L1 experience 'immutably' changes some of the speech perception mechanisms. Whilst early experience/exposure to a particular language environment has an important role to play in restricting initial perceptual capabilities in early childhood, such linguistic restrictions in L1 do not completely inactivate unused innate perceptual mechanisms. In other words, already-established adult perceptions of speech can be modified by later linguistic experience in adulthood.

With regard to apparent discrepancies between results from VOT and /l/-/r/ training studies, i.e. why the modification of perception of the /l/-/r/ contrasts by Japanese adults appears to be more difficult than modification of VOT perception by English-speaking adults, Strange and Dittmann (1984) suggest three possible reasons: 1) L1 allophone-related experience, 2) intrinsic difficulty of the phonetic variation based on language universals, and 3) complexity of acoustic parameters.

2.2 Decline and age

A study by Werker and Tees (1983) investigated non-native speech perception (two Hindi speech contrasts) across childhood (4, 8 and 12 years), in order to determine if the decline between infancy and adulthood in non-native perceptual abilities occurs around puberty, as suggested by Lenneberg (1967). Their results show that the decline is evident by 4 years of age, suggesting that important reorganisations in linguistic perceptual abilities occur in early childhood. Furthermore, they report the relative recovery of discrimination by age 8 for the VOT contrast. This confounds simple maturational explanations, suggesting that non-native speech discrimination does not decline in a gradual linear fashion across development. To explain this, they note that 8 and 12 year olds are more capable of adopting task-specific perceptual strategy, whereas 4 year olds are 'simply rigid rule followers (as has been shown in other cognitive tasks, cf. Kogan 1974)... (Werker and Tees 1983: 285)'.

2.3 The nature of decline

Werker and Tees (1984) provide evidence that age-related decline and modification in non-native speech perception represents a shift in attentional focus/cognitive reorganisation of perceptual processes, not sensory-neural loss. Werker and Tees (1984) and Werker and Logan (1985) found that adult English speakers' performance in discriminating the non-native phonetic distinctions was greater than predicted by chance

when using very sensitive procedures. It was also shown that the adults could not use their sensory capacity in speech perception tasks which required the categorising of full syllables or memory demand, i.e. the ability to discriminate non-native distinctions remained at an acoustic level, and had not been completely lost. Werker (1986) summarised these findings, stating that

> It thus appears that initial phonetically relevant sensitivities are maintained and reorganised into functional phonologic categories if the language to which a child is exposed uses those phones to contrast meaning. The initial phonetically relevant sensitivities that are not exercised do not disappear; rather they become inaccessible for use in a language processing context.
>
> (Werker 1986:142)

2.4 The effect of multilingualism

Research on the effect of multilingual linguistic experience has investigated whether experience with specific stimuli is necessary to maintain/acquire the ability to perceive such specific stimuli or if broadened/non-relevant experience facilitates perception of stimuli in general.

Tees and Werker (1984) conducted a study with three groups of English-speaking adults: L2 learners who had been studying Hindi for 1 year, L2 learners who had been studying Hindi for over 5 years, and those who had early experience of hearing Hindi but no further exposure: these three groups were compared on their ability to distinguish the Hindi voicing (voiceless aspirated/breathy-voiced) and the Hindi place-of-articulation (retroflex/dental) contrast. The group of monolingual English speakers who were exposed to Hindi during the first 1.5-2 years of life (but who could neither speak nor understand more than a few words) could discriminate both Hindi contrasts as adults, suggesting that 'specific' linguistic experience early in life may contribute to maintaining the ability to discriminate phonetic distinctions, even when they have lost the ability to speak or understand Hindi.

The results from Tees and Werker's study provided the hypothesis that, if the developmental reorganisation in speech perception is mediated by specific rather than general linguistic experience, non-relevant multilingualism should not facilitate non-native perception. To test this hypothesis, Werker (1986) compared multilingual adult subjects with monolingual English-speaking adults on their ability to discriminate phonetic distinctions not used in (any of) their native language(s): the Hindi retroflex/dental (syllable) distinction and the Thompson glottalised velar/glottalised uvular (syllable) distinction. Her results have verified that broadened, non-specific linguistic experience does not contribute to increased 'perceptual flexibility'.

Apart from the above findings, Tees and Werker's study also demonstrated that the monolingual English students with no exposure to Hindi performed better for the voicing contrast than for the place-of-articulation contrast. Furthermore, they report that, whilst a short term intensive training resulted in an improved performance of adults in discriminating the voicing contrast, this was not the case for the place-of-articulation. With regard to this point, Tees and Werker proposed the allophone-related 'stimulation history' of individuals in L1.

> Consequently, we believe that the critical difference between our two contrasts involves the stimulation history of our subjects. The evidence in the case of our key contrast involving a place-of-articulation (retroflex) which is seldom used even allophonically in English supports the idea that lack of stimulation did have a significant impact on our subjects' ability to categorise multiple natural exemplars of this Hindi contrast.
>
> (Tees and Werker 1984:588)

This suggests that some contrasts are easier to acquire than others, depending on one's allophone-related experience in L1. With regard to this, the effect of broadened/non-relevant linguistic experience tested in Werker's (1986) study was only limited to the perception of 'novel' contrasts (the two place-of-articulation Hindi contrasts), and thus the allophone-related stimulation history of the subjects would have been minimal. This leads to the more specific conclusion that broadened linguistic experience will not facilitate cross-language phonetic sensitivity towards 'novel' speech contrasts.

This begs the question, which was not addressed in their study, of the effect of multilingual experience on the perception of less 'novel' (or less alien) non-native contrasts, i.e. the investigation of whether broadened/non-relevant linguistic experience facilitates discrimination of non-native sounds that are not distinctive, but which do occur as allophones in certain contexts. This would provide a situation in which the required auditory experience has already been provided (unconsciously) as allophonic variants, with the subjects in a non-contrastive context.

3. Phonetic background for Japanese durational contrasts

Japanese mora sounds /Q/, /N/, /H/ represent one of the best known examples of learning difficulties for L2 learners of Japanese. Thus, the duration of geminate consonants/vowels in Japanese has often been discussed, not only by linguists but also by teachers/learners of Japanese for theoretical and pedagogical reasons. Native Japanese speakers may still detect some kind of 'foreignness' in the speech production of experienced speakers of Japanese, due to their imperfect timing of the geminate vs. single contrasts.

The Japanese geminate consonants consist of sound sequences of two identical sounds (e.g. [pp], [tt], [kk]. The mora obstruent /Q/ is realised as the first part of such geminate consonants and its phonetic realisation is conditioned by the subsequent voiceless consonant. The presence/absence of the mora /Q/ has a phonemic distinction, e.g. /i'ken/ 'opinion' and /i'kken/ 'one house'. In other words, phonemic contrast is realised by a difference in duration of the stop gap, preceding the plosion, although the duration of the Japanese mora sounds itself has been controversial in studies in phonetics/phonology (e.g. Han 1962, Homma 1981, Fukui 1978, Beckman 1982, Han 1992).

Japanese is a language which utilises duration to distinguish meaning. Examples of such variations exist in other languages: long vowels contrasts with short ones in Arabic and long consonants contrast with short ones in Italian. In Japanese and Italian, long consonants occur within a morpheme boundary, whilst in English long consonants exist only across word- or morpheme-boundary as in "white tie" and "unknown" respectively.

This area of contrast between Japanese and English provides the experimental conditions whereby English native speakers have already had experience with long consonants in a non-contrastive context in L1. Thus, perception research with a multilingual group consisting of multilingual speakers of Italian-English (Italian uses short/long consonants phonemically), English-Arabic (Arabic uses long and short vowels phonemically), German-English and French-English, would enable us to see the effect of both specific and non-relevant multilingual experience on the perception of the Japanese durational (short/long) contrast. This less 'novel' contrast in Japanese may be easier for English-speaking learners to recover than, for example, the Hindi 'novel' contrasts tested in the previous studies.

4. Method

The present study was designed in order to investigate the effect of multilingual linguistic experience on the perception of the Japanese durational contrast, which is not strictly 'novel' to English-speaking learners of Japanese, as geminate consonants do occur across word and morpheme-boundaries in English. To examine the effect of both specific and broadened/non-specific multilingual language experience on the perception of the less 'novel' non-native contrast, perceptual identification and AXB discrimination data from the different types of multilingual adult learners of Japanese was compared to those from English-speaking monolingual counterparts. Likewise, the data within the multilingual group was also compared.

The experiment focused on L2 learners' judgment on durational differences between a minimal pair of words /iken/ and /ikken/, which only differed in their stop gap duration. Such data represented perceptual categories for the presence/absence of the Japanese mora sound Q.

4.1 Subjects

10 learners of Japanese at the elementary level, with normal hearing, were tested in the present experiment: 5 multilingual and 5 monolingual subjects. The multilingual group consisted of bilingual (except one who is trilingual) subjects, who were learning Japanese. These multilingual subjects have learned/acquired their L2/L3 as adults and have achieved almost native-like fluency and command of those languages. On the other hand, the monolingual group consisted of all English-speaking subjects, who claimed to 'speak' no other languages (Table 1).

When the testing was conducted, all the subjects had received a nearly identical quantity of Japanese language classroom instruction from a native Japanese teacher for one academic year at British tertiary institutions (7 out of 10 subjects were my own students). Prior to the testing, a language experience questionnaire was conducted from which it emerged that they had the same level of Japanese language experience: they had never been to Japan and their interaction with native Japanese speakers and use of Japanese was limited to the classroom only.

Table 1: Subject descriptions

Multi-lingual subjects	L1	L2/L3	Mono-lingual subjects	L1
S1	Italian	English	S6	English
S2	French	English	S7	English
S3	German	English	S8	English
S4	English	Italian	S9	English
S5	English	Spanish Arabic	S10	English

4.2 Materials and stimuli

Three male native speakers of Tokyo Japanese were asked to read 34 minimal pairs of works with regard to /Q/, /N/ and /R/, in a carrier sentence "Sorewa ____ desu." ("It is ____") twice, and individual recordings were conducted with high quality recording facilities in a sound-proofed recording studio. Amongst all the recordings, the utterances carrying the minimal pair of /iken/ and /ikken/ which were read by a thirty-nine year old male speaker, were chosen as his utterances sounded the most natural and fluent in all respects.

These utterances were subsequently digitised on a SunSPARCstation1, with the sampling rate of 16kHz. Quantificat: was carried out by means of an analog-to-digital (A/D) and digital-to-analog (D/A) converter [2]. Sampling was done at a 16-bit resolution using an Ariel S32C Model 656 linked to an Ariel S32C digital signal processing board. At the same time, the Proport was equipped with digital anti-aliasing filters.

Following this, the decision was made to lengthen the utterance /Sorewa iken desu/ by inserting the duration of silence (i.e. the stop gap) preceding the plosion, using the software, the waves+/ESPS (Entropic Signal Processing Software). In this way, from the original utterance 10 synthetic sentence-stimuli were generated by inserting the duration of 10 msec. silence increments along the durational continuum (Figure 1).

For the identification test, each sentence-stimulus was recorded singly on a cassette tape recorder (MARANTS stereo cassette recorder cp430), with 10 seconds, which were inserted after each of 10 separate randomisations of 10 stimuli. Thus, the identification test consisted of 10 blocks and each block presented a different order of the same 10 stimuli. In total, 100 (10 X 10) stimuli were presented to the subjects.

The AXB test consisted of 5 blocks and one block consisted of 14 trials which were randomised within each block. 14 trials represented 2 AXB orders X 7 possible pairings of stimuli differing by 3 steps along the durational continuum (i.e. 1-4, 2-5, 3-6, ..., 7-10). The AXB test was recorded by inserting an inter-stimulus interval of 1 second, an inter-trial interval of 3 seconds, and inter-block interval of 10 seconds.

4.3 Procedures

The recorded stimuli were presented to the subjects through headphones, at a comfortable listening level using a language laboratory system (Tandberg, TCR 5600)

Prior to the testing, the subjects were instructed orally and a block of stimuli was presented before each test, for the purpose of familiarisation, so that the subjects were informed about what to do in each test. The subjects were asked to indicate their duration judgments by marking 'iken' or 'ikken' on an answer sheet in the first identification test and in the following AXB test, they were asked to indicate whether the second stimuli (X) matched the first (A) or the third (B) for every trial of such a triad. They were asked to answer immediately after listening to each trial, and to guess the answers in the case of uncertainty. It took approximately 10 minutes to complete the identification test and 15 minutes the AXB test, including oral instructions and a practice block before each test.

5. Results and discussion

The perceptual data obtained from the identification test and the AXB discrimination test are shown in the following graphs. On the basis of the scores obtained from the identification test, the calculation of predicted AXB discrimination function was computed using the formula: $Pcorr[A, B] = 1/2[1 + (Pa-Pb)^2]$

Figures 2 and 3 show that the overall performance by the multilingual group is superior to the overall perceptual performance by the monolingual learners in both identification and AXB discrimination tests. Having plotted the means for each group (Figures 4,5 and 6), the comparison of the three groups was conducted on three-way Anova by using a model with one grouping factor (Type = 3 levels) and two within-factors (repeated measures) (Function = 2 levels x Stimulus-Pair = 7 levels). The results revealed that all the main effects were significant (Table 2).

Table 2: Three-way Analysis of Variance

Type	F(2, 7)=13.19	p= .004*
Function	F(1, 7)= 6.23	p= .041*
Pair	F(6, 42)= 4.61	p= .001*
Type by Function	F(2, 7)= 3.88	p= .074
Type by Pair	F(12, 42)=1.58	p= .134
Function by Pair	F(6, 42)= 2.03	p= .083
Type by Function by Pair	F(12, 42)= 1.66	p= .112

*p< .05 Function (Obtained vs. Predicted)

The comparisons of data between specific- and non-specific multilingual groups was conducted on three-way Anova by using a model with one grouping factor (Type = 2 levels) and two repeated measures factors (Function = 2 levels x Stimulus-Pair = 7 levels). The Anova results revealed no significant difference in their perceptual performance, as the main effect of Type failed to reach the .05 level of significance, $F(1,3) = 1.63$, $p = .292$. Likewise, three-way Anova was conducted between specific-multilingual and monolingual groups, and between non-specific and monolingual groups. The results indicated the significant effect of Type in both cases, $F(1,6) = 19.70$, $p = .004$ and $F(1,5) = 9.42$, $p = .028$, respectively. Thus, these results support the hypothesis that not only the effect of specific but also the effect of non-relevant experience increases perceptual performance on the Japanese contrast, which is a 'less novel' contrast. The results have revealed more about the effect of non-relevant,

which reported no significant effect on the perception of 'novel' contrasts by non-relevant, broadened multilingual experience.

Secondly, with regard to the perceptual performance by the monolingual group, the overall pattern of data also revealed a categorical shape rather than a continuous trend which straddled the chance level in both tests. This seems to support the hypothesis that durational contrasts are easier to acquire (or recover) for English speakers than the Hindi place-of-articulation for English speakers or the English /l/-/r/ contrast for Japanese speakers. Werker (1986) and MacKain et al. (1981) demonstrated poor performance by their "monolingual" and "not-experienced" groups (respectively) with much less categorical data straddling the chance level. Perhaps, because long consonants occur across word- and morpheme-boundaries in English, the explanations such as allophone-related "stimulation history" in English by the monolingual subjects, the intrinsic difficulty of these contrasts, or the complexity of parameters (in the case of the stop duration contrast, the parameter is the stop duration only) may be plausible for such an apparent difference between these studies in the (not-experienced) monolinguals' perceptual performance.

Finally, the AXB discrimination-obtained data from the monolingual learners also indicated their existing perceptual sensitivities towards phonetic differences in the acoustic cue, performing as well as the specific-multilingual group for the stimulus pair 7-10 (Figure 5). This seems to provide important theoretical implications, with regard to the mechanisms of speech perception. As the findings from infant and cross-language speech perception research suggest, these results may imply that innate perceptual mechanisms which provide us with perceptual sensitivities toward phonetic differences, are not inactivated completely after L1 acquisition, and that adults may still be able to access (or reactivate) their innate perceptual mechanisms to distinguish between phonetic distinctions that are not used in their L1, in adult L2 phonological acquisition.

6. Conclusion

This paper has reported findings from the perceptual identification and discrimination performance of monolingual and two types of multilingual subjects with the same level of Japanese language learning experience. These perceptual experiments were designed to examine the effect of multilingual language learning experience on the perception of a Japanese durational contrast in the context of SLA.

The results substantiate claims in support of both the effect of specific experience and the effect of non-specific experience on increased perceptual performance with less alien or less novel contrasts. This, in its broad sense, conforms with Ben Zeev's study (1977) on bilingualism in the area of cognition. In testing the hypothesis that 'highly bilingual children process syntactic rules with special flexibility' (Ben Zeev 1977: 1009), she found that bilingual subjects are more advanced in the processing of verbal material, in the discrimination of perceptual distinctions. In addition, whilst showing a greater tendency to look for structure in perceptual situations, they were more able to reorganise their perceptions in response to feedback. Thus, she concluded that bilingualism may lead to increased cognitive flexibility.

In the area of speech perception, Werker (1986) demonstrated the effect of specific experience, but not of non-relevant experience on the perception of alien or novel

contrasts. However, it should be noted here that, in her study, the bilingual subjects were different from those tested in the present study, all her subjects having acquired their L2/L3 between the age of 5-8. In addition, all her subjects were not learning the language whose novel contrasts were tested, i.e. the language was not being learned in the context of SLA, unlike the present study. In this respect, these results may not necessarily be compatible with each other.

The results presented in this study lack empirical and theoretical generalisability because of the numbers of subjects tested: the effect of individual differences cannot be neglected in the interpretations of such a small corpus of data from 10 subjects in total. Thus, the findings presented must be regarded as preliminary and need to be verified/falsified by larger empirical investigations. Such research should shed light on the role of linguistic experience in the development/modification of perceptual categories in L2 phonological acquisition.

Acknowledgments

I would like to thank all the students who took part in the experiments for their cooperation. A short version of this paper was presented at the XVIIIth Annual Congress of the Applied Linguistics Association of Australia (ALAA), 1993.

Notes

1. Voice onset time is referred to as the interval occurring between the beginning of the release of air pressure and the onset of regular vocal cord vibration in the articulation of stop consonants, such as /p/–/b/ and /t/–/d/.
2. Analog-to-digital and digital-to-analog are standard terms used to describe the process of converting from cassette sound recordings to the numbers held in the computer, and converting the numbers back to sound so that we can listen to it and/or record it back to a cassette.

References

Aslin R., D. Pisoni, B. Hennessy, and A. Perey. 1981. 'Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience'. Child Development 52: 1135-1145.

Beckman M. 1982. 'Segment duration and the 'mora' in Japanese'. Phonetica 39: 113-135.

Ben-Zeev S. 1977. 'Influence of bilingualism on cognitive strategy and cognitive development'. Child Development 48: 1009-1018.

Chomsky N. 1981. Lectures on Government and Binding. Dordrecht: Foris Publications.

Eimas P. 1985. 'The perception of speech in early infancy'. Scientific American 252: 34-40.

Eimas P., E. Siqueland, P. Jusczyk, and J. Vigorito. 1971. 'Speech perception in infants'. Science 171: 303-306.

Eimas P., E. Siqueland, P. Jusczyk, and J. Vigorito. 1971. 'Speech perception in infants'. Science 171: 303-306.

Fukui S. 1978. 'Perception for the Japanese stop consonants with reduced extended durations'. Bulletin of the Phonetic Society of Japan 159: 9-12.

Gillette S. 1980. 'Contextual variation in the perception of L and R by Japanese and Korean speakers'. Minnesota Papers in Linguistics and the Philosophy of Language 6: 59-72.

Goto H. 1971. 'Auditory perception by normal Japanese adults of the sounds "L" and "R"'. Neuropsychologia 9: 317-323.

Han M. 1962. 'The feature of duration in Japanese'. Onsei no Kenkyuu 10: 65-80.

Han M. 1992. 'The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers'. Phonetica 49: 102-127.

Homma Y. 1981. 'Durational relationship between Japanese stops and vowels'. Journal of Phonetics 9: 273-281.

Kogan N. 1974. Cognitive Styles in Infancy and Early Childhood. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lenneberg E. 1967. Biological Foundations of Language. New York: Wiley.

Logan J., S. Lively, and D. Pisoni. 1991. 'Training Japanese listeners to identify English /r/ and /l/: A first report'. Journal of Acoustic Society of America 89.2: 874-886.

MacKain K., C. Best, and W. Strange. 1981. 'Categorical perception of English /r/ and /l/ by Japanese bilinguals'. Applied Psycholinguistics 2: 369-390.

McClaskey C., D. Pisoni, and T. Carrell. 1983. 'Transfer of training of a new linguistic contrast in voicing'. Perception and Psychophysics 34: 323-330.

Miyawaki K., W. Strange, R. Verbrugge, A. Liberman, J. Jenkins, and O. Fujimura. 1975. 'An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English'. Perception and Psychophysics 18: 331-340.

Mochizuki M. 1981. 'The identification of /r/ and /l/ in natural and synthesised speech'. Journal of Phonetics 9: 283-303.

Pisoni D., R. Aslin, A. Perey, and B. Hennessy. 1982. 'Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants'. Journal of Experimental Psychology: Human Perception and Performances 8: 297-314.

Strange W. and S. Dittmann. 1984. 'Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English'. Perception and Psychophysics 36: 131-145.

BEST COPY AVAILABLE

Tees R. and J. Werker. 1984. 'Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds'. Canadian Journal of Psychology 38: 579-590.

Trehub S. 1976. 'The discrimination of foreign speech contrasts by infants and adults'. Child Development 47: 466-472.

Werker J. 1986. 'The effect of multilingualism on phonetic perceptual flexibility'. Applied Psycholinguistics 7: 141-156.

Werker J., J. Gilbert, K. Humphrey, and R. Tees. 1981. 'Developmental aspects of cross-language speech perception'. Child Development 52: 349-355.

Werker J. and J. Logan. 1985. 'Cross-language evidence for three factors in speech perception'. Perception and Psychophysics 37: 35-44.

Werker J. and R. Tees. 1983. 'Developmental changes across childhood in the perception of non-native speech sounds'. Canadian Journal of Psychology 37/2: 278-286.

Werker J. and R. Tees. 1984. 'Phonemic and phonetic factors in adult cross-language speech perception'. Journal of Acoustic Society of America 75: 1866-1878.

Figure 1: Stop Duration of Stimuli 1 to 10

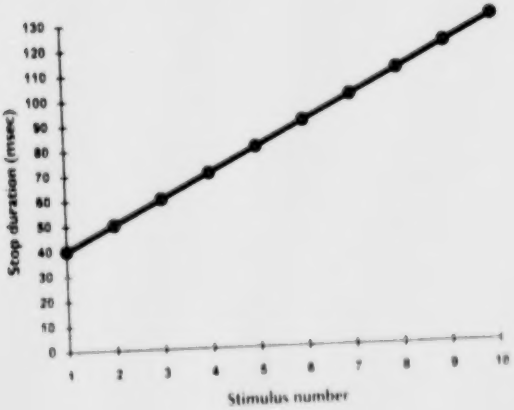


Figure 2: Identification Scores: Monolingual vs. Multilingual Groups

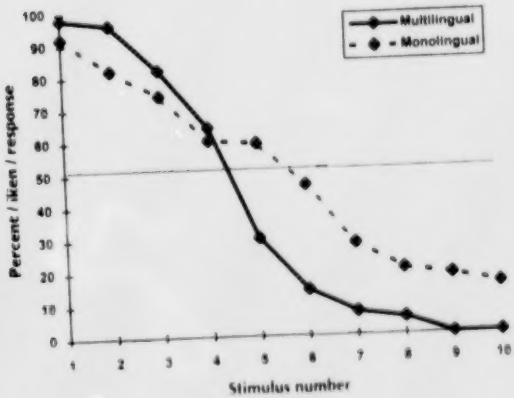


Figure 3: AXB Obtained Scores:
Monolingual vs. Multilingual Groups

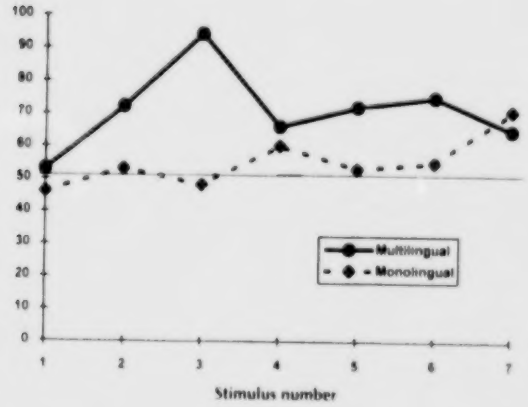


Figure 4: Identification Scores:
Monolingual vs. Two Types of Multilingual Groups

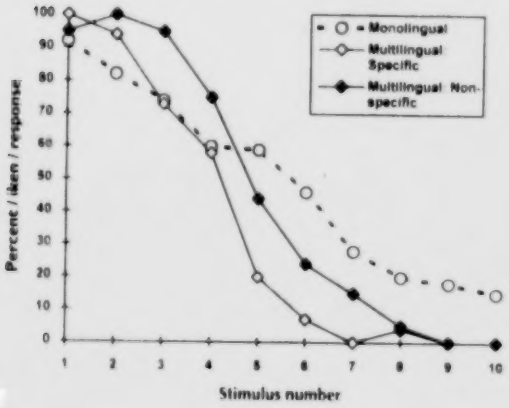


Figure 5: AXB Obtained Scores:
Monolingual vs. Two Types of Multilingual Groups

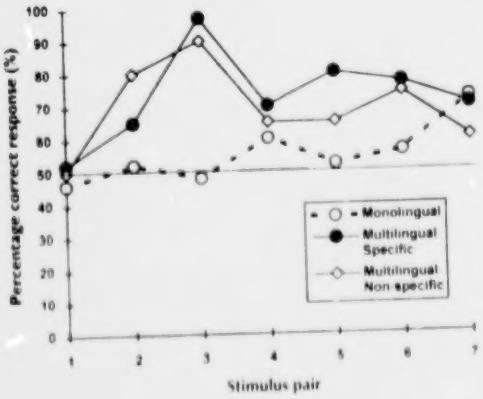
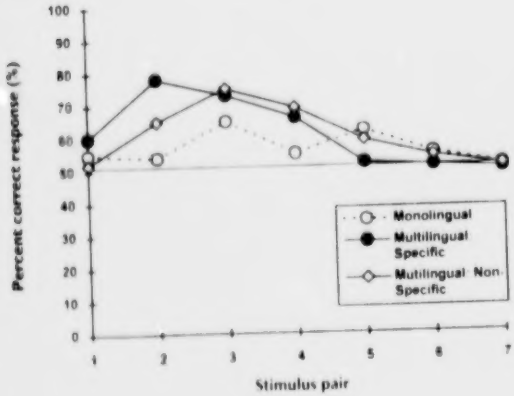


Figure 6: AXB Predicted Scores:
Monolingual vs. Two Types of Multilingual Groups



PEER OBSERVATION AND POST-LESSON DISCUSSION

Sheena Davies and Brian Parkinson (IALS)

Abstract

This paper describes a collaborative project which sought to combine teacher development and illuminative research. Eight teachers, on a General English course, working in pairs, observed each other's lessons (one lesson per teacher) and then held discussions, which were recorded and analysed both qualitatively and quantitatively. Coding systems were used at two stages: by the teachers during the observation (published systems), and by the researchers, who devised a two-dimensional system (topics and 'speech acts') for the quantitative part of the analysis of discussions. Teachers reported that the opportunity to observe was very valuable, and that the published coding systems were useful though at times constraining. Our discussion analysis system, although unreliable on the 'speech-act' dimension, proved helpful in illuminating patterns of interaction.

1. Background

1.1 Types of observation and ways of doing observation

The observation of FL lessons has a long history, but only in the last 10 - 20 years has it "come of age", with a proliferation of observation systems, articles and textbooks.

We can distinguish four main types of observation, according to observer identity and purpose:-

(i) by "experts", as part of teacher "training" and teacher evaluation. This type occurs frequently in pre-service programmes, and is also done by inspectors, head teachers etc.

(ii) by researchers, as part of an attempt to describe and understand classroom events. This may be as part of "pure" research (e.g. Mitchell et al. 1981) or curriculum evaluation (e.g. Parkinson et al. 1981).

(iii) by trainees, observing experienced teachers or each other, presumably in order to learn, crudely speaking, how and how not to teach.

(iv) by practising teachers, observing each other or peers - see Section 1.2.

We can also distinguish three main ways of doing observation, although these distinctions are not clear cut:-

(a) Unstructured: the observer simply makes a note on what he/she considers salient - sometimes no more than a global impression of the lesson. This type has often been associated with (i) above, and is also very common in (iii). It may indicate lack of thought about observation, or reliance on unarticulated professional experience. The rare cases of genuine ethnographic observation can also be included here, though very different in other ways.

(b) Semi-structured/high inference: the observer has a list of questions to answer about the lesson, and sometimes also boxes to fill in, but these tend to require global decisions and exercise of considerable judgement. This way of doing observation is sometimes criticized for unclear criteria and low reliability, but may be justified when observers (and users of the information) have similar expertise and values.

This way of doing observation has been used for all of types (i) to (iv) above. As an example, we give an extract from a checklist for Practical Tests formerly used in the Diploma for Overseas Teachers of English run by the Royal Society of Arts, reproduced in Malamah-Thomas (1987).

PERSONAL QUALITIES	GRADE	COMMENTS
Personality - Presence, general style		
Ability to establish rapport		
Voice: Audibility, ability to project		

PREPARATION	GRADE	COMMENTS
Lesson plan, balance and variety		
Clarity, limitation and specification of aim		
Suitability of materials and methods for level and type of class		

(c) Structured/low-inference: the observer's record of the lesson is compiled according to a detailed rubric or 'coding system' with a pre set list of categories and guidelines on how often to make a coding. Usually but not always, the category requires less inference and judgement than in (b) above. Examples of coding systems are given later in the paper.

This way of doing observation has been common in type (iii), especially the FIAC (Flanders 1960) and Flint (Moskowitz 1971) systems, and even more so for type (ii) (examples later - see Chaudron 1988 for a review.)

1.2 Peer Observation

From our experience, informal contacts and reading it seems that in most or all kinds of teaching (countries, subjects, class types etc.), peer observation is much less common than observation by someone in 'authority' [(i) above] or in training [(iii) above]. Many writers on teaching mention peer observation in passing as a sound idea, but in practice time is short and other priorities, and perhaps embarrassment, interfere. Especially rare is peer observation which is systematic in any sense - either

in that it uses a coding system, or even in that there are definite schedules and objectives.

An important exception is the work of Fanselow, who has stressed the value of peer observation for helping teachers in 'generating and exploring alternatives', and has advocated his own FOCUS system for this (Fanselow 1977). In our experience teachers exposed to Fanselow's work are impressed by his general perspective, non-dogmatic and exploratory, but the FOCUS system itself is found complicated and daunting, and other systems, originally designed for research purposes, are preferred. It seems that the FOCUS system has had fairly extensive use in the author's own circles, and perhaps elsewhere in the USA, but the extent of its wider use and usefulness is unclear.

There have, however, been several recent articles stressing the value of peer observation, e.g. Richards and Lockhart (1990), Lockhart (1991), and teacher development through peer discussion e.g. Edge (1992) and Underhill (1992). These are useful in that they provide different perspectives on the topic but only Richards and Lockhart discuss empirical work and there is no detailed description of their findings. Similarly, though the first two articles both mention lesson coding systems - Lockhart says that after deciding what to look for teachers "can either create or choose an instrument which best codes this behaviour" - details of systems and their operation are absent.

The article by Edge describes a framework for peer discussion, this framework, containing nine categories (adapted from Egan) - Attending, Reflecting, Focusing, Thematising, Challenging, Disclosing, Goal-setting, Trailing, Planning - describes and defines the style of interaction and is part of a particular approach which Edge calls "Co-operative Development". Though it was designed for a differently constrained peer discussion, we found it an original and useful, if indirect, input to our work.

2. The Context Of Our Research

The research took place in 1992 on General English classes at the Institute for Applied Language Studies. The General English course is a full time course of 20 hours per week and the students are, in the main, young adults of various nationalities who come for a full term of 11 weeks. The majority study English to improve employment opportunities, a minority to prepare for post-graduate qualifications. Classes at different times of the day focus on the development of different skills.

The teachers are all well-qualified professionals, and all had had previous experience of being observed and observing others. Three had briefly encountered coding systems some years earlier on a master's course in applied linguistics; beyond this, however, none had ever used particular coding systems and no structured system of peer observation has, to date, been set up at IALS.

3. The Experiment

3.1 Objectives

These were stated as follows in the research proposal:

1. To provide teachers with an opportunity to learn by observing other teachers' lessons, receiving comments on their own and discussing issues arising, supported if required by classroom observation literature (e.g. Allwright & Bailey 1991).

2. To inform the academic community on the outcome of this process (research paper) with particular attention to:

(i) the terms in which teachers conceptualise their own and others' lessons
(ii) what is perceived as different
(iii) what is perceived as surprising
(iv) what attracts positive, negative and neutral comment
(v) what use, if any, is made of classroom observation instruments or other help provided
(vi) in what ways, if at all, teachers would like to continue the peer education and self-education process

3.2 General Procedure

Eight teachers - all those working on the GE course at the time - were invited to participate in the project, on a voluntary basis, and all agreed. They were put into four pairs, and each member of the pair observed one lesson by the other member for at least one hour (lessons last 90 or 100 minutes). The teachers then had a post-lesson discussion in two parts, one for each lesson, each part to last approximately 30 minutes. The researchers were not present at either lessons or discussion, except in one case where a researcher (SD) was one of the teachers. The lessons were not recorded but the discussions were, and the research was conducted on the understanding that discussions, not lessons, were the main focus of investigation. It was also stressed that the research was non-evaluative in a double sense: observers should not 'judge' the lessons, and researchers would not judge the discussion comments; the purpose was mainly one of professional development exploring whatever issues were of interest.

Tapes of the discussions were transcribed (by research assistants who were experienced EFL teachers), and these transcripts (checked with the original where necessary) constitute the main part of our data. They are supplemented by two minor data sources:

(i) the completed coding sheets (see Section 4.1 below) used by the observers and
(ii) a post-discussion questionnaire (see Section 4.2 below).

The teacher/observers were given a selection of recognised coding systems, with background information (see Section 3.3), and asked to select one of these before observing and use it during observation. In addition, they were asked to make notes

of anything observed which was not covered by the system but which seemed surprising, interesting, etc. (cf. research questions, of which observers had a copy). To facilitate this, observers were provided with a three-page coding sheet with space for both system-based comments and open-ended comments.

The observers were not expected to use the systems 'properly', i.e. to make exhaustive coding using exact definitions. This would have been impossible without extensive training. Instead, they were asked to use them as a basis for entries which indicated the main patterns of the lesson. It was stressed, however, that something more than a general impression was required, and that sequential, timed coding should be attempted.

In the post-lesson discussion the teacher/observers were asked to discuss the lessons in whatever way they felt useful. This could be, but did not have to be, based partly on the coding sheets.

3.3 Coding Systems Offered and Used

The following is a brief summary of the systems offered for observer use:

(i) 'BIAS' system (Brown 1975). Indicates whether teachers or learners are speaking, and whether they are lecturing, questioning, responding etc.
(ii) Bowers' system (Bowers 1980). Looks at social functions of classroom language - presenting, directing, organising, eliciting, evaluating, responding, sociating.
(iii) 'COLT' system (Frohlich et al 1985). A high-inference system judging presence of communicative features in interaction.
(iv) Allwright's error treatment system (Allwright 1988). How teacher treats errors.
(v) Chaudron's system (Chaudron 1977). As (iv) but more complex.
(vi) Embryonic system (Long et al. 1976). A list of 17 'pedagogical moves' e.g. initiates discussion, summarizes, provides example; 13 'social skills', e.g. interrupts, contradicts, encourages, jokes; 14 'rhetorical acts', e.g. predicts, hypothesizes, deduces, negates.
(vii) Pica and Doughty system (1985). For the analysis of group-work, and how students check mutual understanding.
(viii) Allwright's turn-taking system (1980) - Taking turns, 'stealing' turns, offering the floor to other learners and so on.

Only the first three of these systems were actually used.

3.4 Data Analysis

This focussed almost entirely on the discussion (other minor sources - coding sheets and questionnaires - are briefly discussed in the results section). Analysis was mainly from the transcripts, but the original tapes were listened to where necessary.

To analyse the transcripts, we devised a two-dimensional coding system, covering (i) the types of topics discussed and (ii) the ways in which teachers interacted, and what we perceived as the underlying speech acts. Our categories were largely 'post-hoc', i.e. created to cover what we found in the transcripts, but we attempted to keep in

mind our research questions and make only those distinctions necessary to answer these.

The system was devised jointly by both researchers, and several joint codings were attempted in order first to improve and then to measure inter-coder reliability. Due to the small amount of data, a rigorous reliability study was not possible but, after coding, a final inter-coder reliability check was made and the results are reported in Section 4.3.1 below.

The categories used, with examples, were as follows:

A. Topics

Number	Topic	Example
1	Facts about students e.g. nationality, level, age, 'history'	"He's been here for ages and he's going to be here for even further ages"
2	Observation system used by coder	"That's the, em the one that is supposed to measure the social functions of the language that goes on"
3	Expectations	"Had you got any preconceived ideas about what would be happening?"
4	Relation between expectations and events	"No well I suppose I imagined probably about what I saw, I mean some input and some practice and some real communication"
5	The observed lesson (general)	"So that's why they had this sort of check list they were using"
5.1	The observed lesson (learner behaviour)	"He was making a face [...] he wasn't seeking confirmation"
5.2	The observed lesson (teacher behaviour)	"And at one point you sort of broke off everything and said 'You look puzzled'"
6	Other lessons by same teacher (general)	"We had been doing quite a lot of work beforehand on discussion techniques"
6.1	Other lessons by same teacher (learner behaviour)	"He doesn't often do that actually"
6.2	Other lessons by same teacher (teacher behaviour)	"In that kind of situation I don't, er, if they ask me a question I just turn away"
7	Other lessons by observer (general)	Similar to 5 and 6
7.1	Other lessons by observer (learner behaviour)	Similar to 5.1 and 6.2
7.2	Other lessons by observer (teacher behaviour)	Similar to 5.2 and 6.2
8	Other lessons or lessons in general (general)	Similar to 5 and 6
8.1	Other lessons or lessons in general (learner behaviour)	Similar to 5.1 and 6.1
8.2	Other lessons or lessons in general (teacher behaviour)	Similar to 5.2 and 6.2
9	Linguistic theories/concepts	"It's amazing how much, well er, I won't say students, I mean anyone can, how much you can read without actually taking any of it in."
10	Materials and syllabus	"Was that something that came up in the textbook?"
11.1	Personal feelings (general)	"I was genuinely surprised"
11.2	Feelings about being observed	"... when you've sort of er being watched by your peers as it were, you do feel a certain, that you are being judged [...] it is a bit nerve racking..."
11.3	Feelings about discussing lesson	"I don't want to sort of make any evaluative comments or judgments on the thing"
12	The English language	"But I mean what is the actual dictionary definition of 'authoritative'?"

BEST COPY AVAILABLE

13.1	The project in general	"Well it seems to me that to get the most value from anything like this, viewing each other once is just nothing like enough."
13.2	Observation in general	"I wonder how far what one learns from observing is actually useful when you're actually teaching. Maybe it is in the planning stage."
14	Other	What did you say?

B. Interaction patterns/underlying "speech acts"

Label	"Speech Acts"	Examples
BC	Back channelling	"Ah, I see. Right, right."
SY	Sympathising	"This is a big problem isn't it, that everybody has."
SP	Speculating	"Maybe she only talked a lot the second time because she was [playing another role]."
QN	Question (neutral)	"Was it difficult to try to use it?"
QC	Question (challenge)	"But why did me being there stop you doing it spontaneou[sly]?"
INF	Informing	"After you left they started discussing again."
OP	Opinionating	"I think X is very much a sort of actor anyway."
EV+	Evaluation (positive)	"It was good that it was slipping into real discussion."
EV-	Evaluation (negative)	(i) "[My] board work was pretty awful" (ii) "The only thing was, the only thing was, that it was, they, the particular constructions with the particular prompts did seem to be causing some rather odd constructions."
IE	Inviting evaluation	"Was there anything else that surprised you, that you wanted to ask me why you did something? Cos it's very useful for me as well to get sort of, a reaction of some kind."
A	Adverting or semi phatic	"You reminded them about erm what they had done the previous week."
J	Justifying	"I mean the point of that exercise was to see whether they had actually got their heads round the, erm, the distinction."
SG	Suggesting	"If they'd had a piece of paper with it written down it probably would have been easier."
AG	Agreeing	"Yes, Anna said that a couple of times, yes."
Other	Other	"Oh, maybe."

N.B. These "speech acts" must be interpreted in their interaction with different topic categories, and with who (teacher or observer) is speaking. For example, commonsense suggests that evaluating one's own lesson, one's interlocutor's lesson and a third party's lesson are three very different kinds of speech acts. The grand total for each "speech act" may thus sometimes be less illuminating than totals for individual combinations of topic and "speech act".

4. Results

4.1 Information from Lesson Coding Sheets

Six of the eight teachers handed in their coding sheets. (They were not obliged to do so since the focus was on the post-lesson discussion.)

When analysing the data, we asked ourselves the following questions:

- how long did they keep on coding?
- did they use their own categories or those of the 'system'?
- what was the balance between the 'system' observations and the open-ended observations?
- what type of comment/information was in the open-ended column?

- did they use the third checksheet (for general comments)?

All the teachers coded the lesson events for the full hour, following the time segments. The use of the categories of their chosen systems varied from full use of all the categories (by two teachers using the BIAS system and one using BOWERS) to partial use of the categories (by two using COLT and one BIAS). Most teachers gave some illustrations or explanations of these categories in the system column. For example: *"Sociating - rearranging bodies"* (BOWERS); *"S [= Silence] because focussing their minds & individually formulating questions"* (BIAS).

In the open-ended column, most comments were descriptions of lesson events - a combination of more illustrations of the system categories and descriptions of other events focussing on teacher or student behaviour, or examples of language used, such as, *"Qs on OHT on topics of film"*, *"E demonstrate swagger by walking"*. One sheet also included several positive evaluative comments: *"most Ss participating well"*; *"T good at involving any S who is reticent left out"*.

The balance of observations between the two columns seemed fairly even, and in most cases there was nothing substantially different between the two, apart from one coding sheet in which the open-ended column consisted mainly of diagrams of the teacher-student interaction. This reflected the personal interest of the teacher concerned, an interest she elaborated on in both the post-lesson discussion and the general checksheet 3.

Checksheet 3 gave observers the opportunity to make any further comments and was submitted by only four teachers. They each used Sheet 3 for further descriptions of lesson events, but two also made comments and posed questions about the system and what they were trying to do. For example, *"Is communication genuine if T clarifies language for S?"*; *"Difficult to concentrate on different levels of communication e.g. interpersonal, formal, plenary"*; *"Prediction of a lesson format v. difficult if not impossible"*. These issues were later raised in the post-lesson discussion between the teacher and the observer.

4.2 Information from Questionnaires

Tutors were asked to fill in a questionnaire to supplement the data from the discussion. The questions were open-ended (see below) and anonymity was guaranteed.

1. Why did you choose the observation system(s) you did? Was it/were they a help or a hindrance?
2. What, if anything, did you find useful or valuable about
   (a) observing another's class?
   (b) being observed?
   (c) discussing the lessons?
3. Did you feel any constraints in the post-observation discussion?
4. How, if at all, would you like to follow this up?

The responses, though varied, show some measure of agreement. Not surprisingly perhaps, four out of the eight tutors stated that they had chosen the system because it

reflected an area of interest, while the others selected a system for its apparent manageability for real-time coding and its "degree of teacher-friendliness".
"I chose a system that seemed best suited to look at teacher-student interaction" (BOWERS); *"It seemed reasonable to handle within the constraints of timing and is an area of personal interest."* (BIAS)

With regard to its usefulness, there were only two unqualified answers - one found it useless, the other a positive help by providing a systematic framework. The remaining tutors found it of limited usefulness - a help initially by giving a guideline or focus but then becoming a hindrance because it was too restricting or difficult to use, or inappropriate for the actual classroom events. *"A good guideline, but perhaps too narrow."*; *"On the whole it was a help but sometimes it was difficult to observe other useful and interesting events because of the focus of the observation system."*

All the tutors, however, were positive about observing another's class and there was strong agreement in the reasons given, such as: *"interesting and informative to watch how someone else deals with a topic ..."*; *"makes you review your own teaching methods"*; *"it is always useful to pick up new ideas from other teachers"*

The comments on being observed echoed the comments above, for example: *"having another person's opinion about a problem"*; *"makes you think a little more about what you do and say, and how much you say"* and also included the opinion that *"As I believe in team teaching as a useful method of teaching certain types of classes and students, I think it is helpful to be observed and to be able to feel comfortable and not in any way inhibited by the presence of another teacher."*

Similarly, the post-lesson discussion attracted positive comment - again, mainly echoing the points made above but two teachers made similar observations about interpretation and perceptions: *"interesting to see whether what the teacher felt had been important/salient corresponded to what the observer felt"*; *"I found it interesting, and quite surprising sometimes, how an observer interpreted what I did"*

Only one person found the experience of being observed and the post-observation discussion nerve-wracking, nor did this teacher like the discussion immediately after the lesson, preferring (in retrospect) to have had more time to reflect and be more analytical.

There was a high level of unanimity among the responses to the questions on feelings of constraint in the post-observation discussion. Five tutors commented that they would not want to say, or tried to avoid saying, anything critical about their colleague's lesson because it was a peer situation. This does not mean that they felt there had been something negative to say, only that they would not have wanted to say it if there had: *"If I really didn't like something I saw or had a negative criticism (which was not the case in the one observation) I feel that this set-up would not encourage me to say anything."*; *"I wouldn't want to say anything negative (even if I had something to say) as it was a peer situation and as such should be supportive and non-judgemental."*

This conforms with the emphasis on the non-evaluative nature of the research project in general and the lesson observation in particular (see 3.2 above) but it is interesting that such a feeling should be perceived as a constraint. Only one person implied that

there were negative points made about the observed lesson. Other constraints mentioned were the artificiality of talking into the microphone (initially), and talking in a more 'formal' way rather than a chat over a cup of coffee.

With regard to possible follow-up, the points made tended to fall into two main categories:

(a) introduce a system of frequent observation as routine teacher development, not as a research project (five teachers mentioned this)

(b) if the project is repeated, to have a more focussed system on topics agreed on by participating teachers (two teachers)

4.3 Introduction to Discussion Coding Tables

The following tables give the overall coding totals for observers and for teachers. They show the number of occurrences in each category, and each combination of "topic" and speech act category.

Table 1
Observer Moves - Grand Total

	BC	SY	SP	QN	QC	INF	OP	EV+	EV-	IE	A	CN	I	SG	AG	Other	TOTAL
1 Facts about students	6	2	1	15	4	1	3				3				2		37
2 Coding system	2			1		38	4	1	11				1			1	59
3 Expectations				1		4			1								6
4 Relation expect event				1	1	5			1		1						9
5 Observed lesson (general)	22	5	3	9	2	21	5	28	6	1	16	2	1	3	1	2	127
5.1 Observed lesson (SS)	7	1	5	7	4	28	26	23	7		37		5	2	12		164
5.2 Observed lesson (T)	1	2		8	5	17	9	37	10		31		9	6	1		136
6 T's lesson (general)	10			9	4	2	2										27
6.1 T's lesson (SS)	7		1	5			1										14
6.2 T's lesson (T)	3	1		8			3						1				16
7 O's lesson (general)		1		2		12								1			16
7.1 O's lesson (SS)						1											1
7.2 O's lesson (T)						4											4
8 Other lessons (gen.)	2	7	1	3		6	5				1			1	3		29
8.1 Other lessons (SS)	2		1			3	8				1				3		18
8.2 Other lessons (T)	3	1	2			3	3							1			13
9 Theory			2			4	6										12
10 Material	6		1	9	2		1	1			3						23
11.1 Feelings (general)	1			1		10											12
11.2 Feelings (observed)			4	2		5	7					1			2		21
11.3 Feelings (discussing)						5			1								6

BEST COPY AVAILABLE

12	English	2		1	4	1	3	1				1	1			1		15
13	Project/obs (general)			1			1	4										6
14	Other				2		1									1	2	6
	TOTAL	74	20	23	87	23	174	88	90	37	1	94	4	17	14	26	5	777

Table 2
Teacher Moves - Grand Total

		BC	SY	SP	QN	QC	INF	OP	EV+	EV	IE	A	CN	J	SG	AG	Other	TOTAL
1	Facts about students	2					35	4	2			1				4		48
2	Coding system	6		1	6		1	3		1		2	2	1		1		24
3	Expectations				2		1											3
4	Relation expect/event				2		9		1		1							13
5	Observed lesson (general)	16			6		28	9	8	4	8	16	3	31		8	1	138
5.1	Observed lesson (SS)	3	1	4	5		29	20	13	5	1	17		6		11	1	116
5.2	Observed lesson (T)	10		3	2		7	1	6	4	9	4	9	33	3		1	92
6	T's lesson (general)	1			1		35	5		1	1		1	3		2		50
6.1	T's lesson (SS)			1	1		42	11	3	3		1		1				63
6.2	T's lesson (T)						35		1		1		1	14		1		53
7	O's lesson (general)						2					1						3
7.1	O's lesson (SS)																	0
7.2	O's lesson (T)																	0
8	Other lessons (gen.)	2		1			2	8			1	2		1		3		20
8.1	Other lessons (SS)	1	1				3	9	1	2		1				3		21
8.2	Other lessons (T)			2			3	7					1		1			14
9	Theory	1					2	5								1		9
10	Material	1		1			18	1	1	2		2	2	1				29
11.1	Feelings (general)			1	1	1	6	1					1					11
11.2	Feelings (observed)	2		4		4	4	4								2		20
11.3	Feelings (discussing)																	0
12	English	1	1	1	1		2	3								1		10
13	Project/obs (general)						1	2										3
14	Other	1			3		4	1								1		10
	TOTAL	47	3	19	30	5	269	94	36	22	22	47	20	91	4	38	3	750

4.3.1. Inter-coder reliability

As an inter-coder reliability check, half of one discussion, chosen at random, was coded by both researchers. We chose not to give a single percentage figure or set of figures for reliability, considering this to be meaningless as it could be calculated in many different ways. The statistics seem to show, however, that our 'topic' figures have some inter-subjective status: agreement is as high as could be expected in a small, exploratory study. The "speech act" agreement, on the other hand, is disappointingly low, and means that our figures on this dimension must be treated sceptically, though they can give some general indication of how teachers behaved.

As far as we can tell, our difficulties were not mainly to do with conceptualising the categories: we seemed to agree on what we meant by informing, evaluating etc. - but on interpreting specific utterances. Illocutionary force is often not totally explicit, and it is well known that it may be difficult for third parties, and even for participants, to recognise it with certainty.

4.4 Annotated selection of teacher comments

This section aims to supplement the quantitative information on the discussion with a selection of teacher comments, organised by topic, which seem particularly relevant to the research questions. We have omitted hesitation devices and repetitions as far as possible, except where they seemed important to an understanding of the interaction, but we have not achieved total consistency in this, nor in our indication of minor abridgement. Comments by both observer [O] and the teacher [T] being observed are included.

4.4.1 Comments on use and usefulness of coding system

As already indicated, we asked teachers to be guided by the system rather than to attempt exhaustive coding of every utterance, and most did exactly that. Nonetheless, several offered reasons (in the discussions) for their limited use of their chosen system, often coupled with comments on its adequacy or otherwise:

"I was trying to use ... Brown's Interaction Analysis (BIAS) ..... When I looked through the ones on offer, it seemed to be quite interesting. It was in the area that I wanted to consider. But ... it's very difficult to actually stick to one of these systems exactly, because there aren't enough categories. Or these categories given here are not exactly relevant to the type of lesson you were doing. I would have liked to include more categories because some of these ... sound negative e.g. 'teacher lectures'. To me that has a negative overtone that the teacher's jawing away all the time. Then it says "explains" and "directs": well yes, you did that a few times, on the sheet I've put down "TL" in various situations, but it was very short, succinct and straight to be point ..." [O]

We would not presume to reduce such comments to a one-sentence summary, but the main points are perhaps (i) the "ethnographic" one that pre set systems fail to capture the richness of experience, (ii) mismatch between system and modern methods and (iii) distrust of evaluative labels.

4.4.2 Comments on effect of observation

As usual in studies of this type, the question arose of the effect, if any, of observer presence on teacher behaviour and learner behaviour. Effects on the teacher were only rarely mentioned: *"I'm always a bit nervous though, I think you tend to be, no matter what you say you still think you're being watched and evaluated."* A possible effect on learners was mentioned slightly more often: *"They're usually quite a lively class specially first thing in the morning ... so it might have been the fact that you were here that might have had an influence"* The majority view, however, seemed to be that the effect of observation was small, e.g. *"No, I don't think they were affected. X has been in to watch them for a lesson, and people pop in and out, and I film them ... so I don't think they're bothered too much by other people"*

4.4.3 What attracts positive comments

Although all observers showed a positive and supportive attitude towards the teacher and the observed lesson, they did not find it necessary to make frequent comments of direct praise. Interestingly, many evaluative comments related to students rather than the teacher, although some of these might be interpreted as indirect praise of the teacher: *"They seemed to be working very naturally together"*; *"They really did seem to get on well together, that was nice. I mean there was quite a lot of er, I was quite surprised, there was quite a lot of touching actually between them ... you know, playful slaps on the and that kind of thing"*; *"I mean there was a tremendous enthusiasm there wasn't there, they were really sort of getting in there"*

An example of direct positive comment on the teacher was: *"I felt that the lesson was very carefully crafted."* Less clear as a speech act, but also classified as positive comment on the teacher, was the comment: *"it was quite interesting the way you were anticipating that they would have problems, you'd be much more careful about how you communicated"*

An area in which reassurance was sought in several discussions was that of learner involvement. Teachers said that they, often or on a particular occasion, were unsure if activities went on too long or not long enough, if some students were bored etc., as they were unable to monitor all learners all the time. Observers replied positively in all cases - as far as they could see pacing and involvement were satisfactory, though in some cases it was agreed that differences in learner level etc. meant there was no ideal solution.

4.4.4 What attracts negative comments

It is a general rule of human interaction that negative comments are made far more often about (especially absent) third parties, far less often about the speaker, least often about the addressee, so it is no surprise that most comments in the 'negative evaluation' category referred to students, e.g.: *"A tends to wander off into his own little world quite a lot"*; *"B can be quite aggressive"*; *"There's one or two who don't like being corrected ... seem to lose face a bit"*

It is striking, however, that such comments usually refer to one student, most of the rest to small sub-groups, very few to whole classes. The teachers generally seem to

have positive attitudes to their classes, and even the criticisms of individuals may usually be interpreted as relating to the difficulty of doing certain kinds of work with them rather than as outright hostility.

Self-criticism was only a minor feature of the discussions although not infrequent it was typically brief, and in the nature of an aside: *"My board work was pretty awful."*

Criticism of the teacher is infrequent and almost always mitigated or qualified. We do not suggest, however, that observers were withholding or excessively 'toning down' negative comments for reasons of tact. One can never know, but the evidence (see 4.2) and our impressions suggest that the teachers did feel generally positive about the lessons and about each other. But the difficulty of expressing even minor and constructive criticism needs to be borne in mind.

4.4.5 What is perceived as surprising?

Observers seemed to find the lessons interesting, but rarely surprising. This was perhaps to be expected as there is a culture of cooperation within General English and a lot of informal discussion about what happens in classes. For an exceptional 'surprise', see 4.4.6.5.

4.4.6 Specific issues of methodology and learner behaviour

Under this heading, we consider methodological and similar issues which arose , some in only one discussion, some in most or all discussions.

4.4.6.1 Genuine communication versus practice of forms

The first of these, the proper rôle for 'communication' activities and practice of forms, and indeed how to distinguish between these, was addressed in some form by most of the participants, most extensively when the COLT system had been used:

"... what we're trying to achieve in the afternoon classes, we're trying to promote real communication but we have to give practice activities ... specially advanced level ... they're not like practising one particular function, it's usually some kind of a strategy ... so the fact that you give them a task in which you want them to practise this but the task itself is not too constraining, is that genuine communication or not?"
(and later)
"When you were eliciting things from them, and although you didn't know what their responses would be, is that genuine communication or not?" [O]

Learners in some classes had been given a range of exponents for expressing opinion, taking the floor and related functions, and when they practised these there was sometimes doubt about whether they had to express genuine opinions. *"It was very interesting to try to gauge at different points to what extent they were having a genuine discussion [or] just practising language very consciously and it seemed to slide very much backwards and forwards between the two."* [O]

BEST COPY AVAILABLE

4.4.6.2 Error treatment

This topic arose with similar frequency to the last, and at greater length. Space prohibits a full account of conclusions reached, but we quote from two discussions to give a flavour of the ideas expressed. *"... it was quite interesting to know through the progression of the lesson whether it's certain activities that lend themselves to a certain kind of error correction, in which case it's quite important to plan one's lesson."* [O]

In the following example, an observer reflects how different kinds of class at different times of day (see Section 2) require different attitudes to error: *"I expect it to be more teacher-directed than student-centred which in fact happened and I also expected there to be more error correction, you know, than in a fluency class or something like that ... Error corrections and pronunciation checks and things like that ... so, my preconceptions were justified so it wasn't surprising from that point of view, you know, given the kind of lesson it was."*

4.4.6.3 Teacher as model (articulation)

This is an interesting example of a one-off comment which derives from an important issue on the observer's personal agenda. No criticism seems to be implied, only 'wondering aloud': *"it was quite interesting ... and this is something I always worry about, you were articulating very carefully when you were speaking to the students. I do that ... I just wonder are some teachers doing it, you know as a sort of, saying it in a very fluent native speaker way and not articulating?* [O]

4.4.6.4 Dominance in group work

The observer in one discussion chose to investigate whether particular learners were dominant in group work whilst others said little: this is of course a common area of both system-based and open-ended observation, promoted especially by Allwright. The conclusion was that in this case no-one was obviously dominant, and no-one excluded, but an interesting pattern was noted: *"It was A and B who were doing most of the talking, and then the blonde woman, well her role seemed quite interesting ... she seemed to be slightly also taking on the teacher's role in that she sometimes, I think she once or twice corrected people or provided words .but somebody else was looking for, ... they seemed to ... look to her ... as a sort of linguistic consultant."*

4.4.6.5 Learner independence

One discussion was different from the others in that, in the observed lesson, the teacher had adopted a range of procedures which both teacher and observer perceived as somewhat unusual or individualistic, and which had an explicit rationale of encouraging learner independence. Thus, for example, the class had watched TV programmes, chosen by themselves, aided by their own lists of predictions/questions rather than teacher-devised worksheets, in order to show them a way to benefit from watching normal TV at home: and many questions about vocabulary, even requests for guidance on 'the best way to learn', were turned back for the students to answer themselves. In the discussion, the observer generally praised this approach (*"I love the way you did that"*), but also voiced doubts, e.g. *"I'm just wondering what happens*

if nobody knows". The teacher elaborated the thinking behind the approach, but modestly disclaimed any definitive answers and accepted that the approach would not always work, even giving an example of a 'failure' after the observer had left the lesson.

As mentioned above (4.4.5), not much in the observed lessons appeared to be perceived as surprising, and this discussion seems to be the only case within the discussions of something being described as radically new.

Other teachers too were very conscious of the learner independence issue, and provided examples in their actions as teachers or their suggestions as observers. One teacher, for instance, had invited a student to explain something to another student, and justified this as follows: *"Well I like to do that because, rather than me just saying 'It means this', which in that state I'm becoming the dictionary and I'm trying to wean them off these wretched dictionaries so it's better if they're trying to ..."*

The next example shows the same desire for learner independence, but a less satisfactory reality: *"what's really unfortunate is even if they are in groups, they'll still kind of try and turn round and involve me and sort of call me in you know."*

Finally, the 'learner independence' area provides one of the few examples in the post-lesson discussions of a detailed suggestion from an observer: *"I noticed that when he'd maybe finished he just sort of tended to, he'd obviously, what was his name, A had obviously finished and you think well, he could have been encouraged more to help, or maybe, I don't know, maybe he didn't like helping his friend, you know, and so he maybe, getting him to explain 'Why did you put this?', 'Why did you put that?'"*

4.4.7 Feelings about doing observation

The general idea of observing each other's lessons seemed to be viewed very positively. As regards the use of systematic coding systems, or any kind of systematic peer observation, comments were more cautious - still positive, but aware of the difficulties and the need for more experience. Sample comments: *"I suppose it's very significant that when teachers sit down to talk about a class, the mode that you tend to slip into is evaluative, even if it's very encouragingly, because most often any time you're actually paid to sit down and talk about someone else's class it's because you're examining them or you're helping them towards an exam, or you're seeing how it should be done from somebody who has more experience"* [T]; *"you're switching into observer mode for a particular thing and you're not aware of of all the other things that are going - you're looking for X and there's A to Z going on around you because you're looking for X"* [O]; *"I think as practising chalkface teachers we're very interested in how things work - and classroom management techniques and skills and so on ... it's relevant to this observation thing that we're doing that you tend to get caught up with those pedagogical issues"* [O]

5. Discussion And Conclusions

5.1 Discussion of Transcript Codings: 'Topic' Dimension

The following general features may be noted:-

(i) In all discussions for both speakers the observed lesson (category 5, including 5.1 and 5.2) is, not surprisingly, the main topic.

(ii) Within this category the emphasis varies, some discussions emphasising learner behaviour, some teacher behaviour, some more evenly balanced. On the whole, however, learners were discussed more than teachers. This may reflect both the current climate of opinion within EFL and the perceived usefulness of another perspective on learner interaction.

(iii) The second most consistently frequent category - more than 10 codings in each of six discussions - was category 6 (including 6.1 and 6.2), i.e. what the observed teacher did in other lessons, including parts of the observed lesson before and after observation. The felt need to put observed events in context is unsurprising.

(iv) The only other categories which even approached this frequency - each having more than 10 codings in each of four discussions - were category 8 (including 8.1 and 8.2), i.e. lessons in general, and category 2, the coding system.

(v) Only three other categories ever exceeded the '10 codings' threshold, and these only in one discussion each, reflecting a particular focus of that discussion. These were: The teaching material (category 10); Feelings about being observed (category 11.2); The English language (category 12).

(vi) There was a tendency for similar topics to be discussed in each of a pair of discussions with the same participants.

At this point it may be worth repeating that the ethos of the project was non-evaluative, and we shall not presume to suggest that high or low use of any topic is 'good' or 'bad'. The teachers discussed what they identified as worth discussing, exactly as intended. If one were planning a more extensive and structured peer observation study, however, one might wish to look for ways, over a long series of discussions, of ensuring a wide and systematic topic coverage including both teacher and learner behaviour.

5.2 Discussion of Transcript Codings - 'Speech Act' Dimension

The first point to note here is a blurring of the role distinction which might have been expected: for example, not only the observer but also the teacher sometimes evaluates what happened in a lesson, and not only the teacher but also the observer justifies it. This perhaps indicates a high level of mutual supportiveness with the group, with strong desire to convey feelings of solidarity and emphasize the shared features of experience.

Beyond this, we were struck by the wide variation of speech act frequency in the discussions: the differences between the discussions are far more salient than the similarities. For example, two discussions, with their high frequency on the part of the observer, of questioning and 'adverting or semi-phatic' (i.e. mentioning something known to both speakers), seem to be a very different kind of speech event from other discussions, where these categories are much rarer. (Other untypical figures, such as the high 'back-channelling' rate of one observer, may reflect individual speech styles.)

What is common to all discussions, however, is a fairly high rate of 'evaluating' acts by the observer (sometimes, only late in the discussion and after frequent invitations to evaluate by the teacher.) Most of these comments were on the observed lesson, although other topics (materials, students, other lessons) were sometimes evaluated. All teachers made one or more evaluative comments on their own lessons, though sometimes not many and always fewer than the observer. Evaluations were in general positive, and any negative comments were tentative and qualified.

The frequency of evaluative comments seems to confirm that peer observation is always likely to be a partially evaluative process, and that, although one can mitigate this by emphasizing non-judgmental aspects, one cannot, and perhaps should not want to, eliminate it entirely.

5.3 General Conclusions

Our first, and unsurprising, conclusion is that the peer observation was worth doing. Teachers did appreciate the opportunity, surprisingly rare in most professional lives, to observe another teacher's lesson and to be observed without any context of evaluation or bureaucratic requirement.

Second, and less certain a priori, the coding systems seemed to be of some value. Every professional in the area of systematic observation knows that all existing systems are far from ideal, sometimes difficult to apply - especially to classes taught by modern methods - and not always yielding insights, and the observer comments (see especially 4.4.1) amply confirmed this. Nonetheless, observers did persevere with systems, often filling in coding sheets very fully - perhaps more fully than we had expected - and showing great resourcefulness in taking the systems as a starting point for more open-ended comments on topics of major interest.

Our third conclusion relates to the low inter-coder reliability on our 'speech-act' dimension. It seems either that we are unusually incompetent in recognising illocutionary force, or, more probably, that what is said in discussions of this type is even more polysemic (polypragmatic?) than one would generally suppose. This could be a fruitful area for further research.

Fourthly, despite these uncertainties of pragmatic detail, the general goodwill, enthusiasm, mutual supportiveness and professional commitment of the teachers was very much in evidence, as was the structured and principled nature of the curriculum which they were implementing.

Fifthly, it seems that further peer observation, although not an urgent priority, would be of potential benefit to the course, to those of the eight teachers still working on it and to others who have replaced some of them. As usual in such research, any second round of observation could profitably be made slightly more selective and structured, building on the findings of this pilot study. Post-lesson discussion could be similarly guided, with a wide and systematic topic coverage including both teacher and learner behaviour.

Acknowledgement

We would like to thank the teachers who took part in this project - Wendy Ball, Cathy Benson, Sue Fraser, Marie Gilroy, Liam Rodger, Paul Snowkes and Richard Warner

References

Allwright R. 1980. "Turns, topics and tasks: patterns of participation in language learning and teaching". In D. Larsen-Freeman (ed.) Discourse Analysis in Second Language Research. Rowley, Mass.: Newbury House. 165-187.

Allwright D. 1988. Observation in the Language Classroom. Harlow: Longman.

Allwright R. and K. Bailey. 1991. Focus on the Language Classroom. Cambridge: CUP.

Bowers R. 1980. 'Verbal Behaviour in the Language Teaching Classroom'. PhD Thesis, University of Reading.

Brown G. 1975. Microteaching. London: Methuen.

Chaudron C. 1977. 'A descriptive model of discourse in the corrective treatment of learners' errors'. Language Learning 27: 29-46.

Chaudron C. 1988. Second Language Classrooms. Research on Teaching and Learning. Cambridge: CUP.

Edge J. 1992. 'Co-operative development'. ELT Journal 46.1: 62-70.

Fanselow J. 1977. 'Beyond Rashomon - conceptualizing and describing the teaching act'. TESOL Quarterly 11: 17-39.

Flanders N. 1960. Interaction Analysis in the Classroom, a Manual for Observers. Ann Arbor: University of Michigan.

Frohlich M., N. Spada, and P. Allen. 1985. 'Differences in the communicative orientation of L2 classrooms'. TESOL Quarterly 19: 27-56.

Lockhart C. 1990. 'Co-operative teacher development: new observations on observation.' Perspectives 2: 43-57.

Long M., L. Adams, M. McLean and F. Castaños. 1976. 'Doing things with words - verbal interaction in lockstep and small group classroom situations': in J. Fanselow and R. Crymes (eds.) On TESOL '76. Washington: TESOL: 137-153.

Malamah-Thomas A. 1987. Classroom Interaction. Oxford: OUP.

Mitchell R., B. Parkinson, and R. Johnstone. 1981. The Foreign Language Classroom, an Observational Study. Stirling Educational Monographs No. 9. Stirling: Department of Education, University of Stirling.

53 BEST COPY AVAILABLE

Moskowitz G. 1971. 'Interaction analysis - a new modern language for supervisors' Foreign Language Annals 5: 211-221

Parkinson B., R. Mitchell, and D. McIntyre. 1984. An Independent Evaluation of 'Tour de France'. Stirling Educational Monographs No. 11. Stirling: Department of Education, University of Stirling.

Pica T. and T. Doughty. 1985. Impact and interaction in the communicative classroom: a comparison of teacher-fronted and group activities. In S. Gass, and C. Madden, (eds.) Input and Second Language Acquisition. Rowley, Mass.: Newbury House. 115-132.

Richards J. and C. Lockhart. 1991. 'Teacher development through peer observation' Perspectives 3/2: 1-9.

Underhill A. 1992. 'The role of groups in developing teacher self-awareness' ELT Journal 46, 1: 71-80.

ARE SCORE COMPARISONS ACROSS LANGUAGE PROFICIENCY TEST BATTERIES JUSTIFIED?: AN IELTS - TOEFL COMPARABILITY STUDY.

Ardeshir Geranpayeh (DAL)

Abstract

Many academic institutions in the UK and Australia require their non-native candidates to provide a proof of a certain band score on IELTS or its equivalent score on TOEFL as evidence of English proficiency to pursue a course of study. This study is concerned with whether score comparisons across TOEFL and IELTS are justified. The results reported here suggest that score comparisons across TOEFL and IELTS are possible but institutions should be cautioned about the comparability of the test scores and should allow for possible extraneous factors affecting these scores.

1. Introduction

1.1 Introduction

Hundreds of thousands of individuals throughout the world take various English language proficiency tests each year to demonstrate their proficiency in English as a foreign language. The scores of such tests will then be used by different institutions for screening their candidates for a number of different purposes such as offering employment, advancement in a career, or admission to an educational programme. In most cases, the selection of candidates is affected by the results of these tests. Thus, any variability in the scores of such tests might affect job opportunities or perhaps life chances of individuals. This makes the interpretation of the scores an extremely heavy responsibility.

Test scores could be related to various aspects of proficiency demonstrating the candidates' language ability in different skills, i.e. writing, reading, speaking, or listening in a given language. In the last three decades, numerous methods and test batteries have been developed to measure different aspects of language proficiency of non-native speakers. Depending on the nature of the test population and the purposes to which the test scores are put, the tests presumably differ from one another. In most cases, differences in methods and purposes are considered as evidence for the incomparability of LP tests. Yet, where the statistical evidence is concerned, the tests are validated against one another and their results are compared to show the degree of similarity between the traits they are measuring.

On the other hand, academic institutions are interested only in a clear cut-off point score of, say, 600 on TOEFL or its equivalent 6.5 on IELTS as evidence of their non-native speakers' suitability to pursue a course of study. What does it mean to have a specific score on a test? How can a quantitative value obtained in an hour's

EDINBURGH WORKING PAPERS IN APPLIED LINGUISTICS, number 5 (1994) ISSN 0959-2253

testing period predict the future success of a candidate in following a career? How can different scores obtained in different batteries be equated to one another? These and many more questions have been raised in the literature of language proficiency testing.

This paper is an attempt to clarify one of the relevant issues in comparability studies, that is, whether score comparison across test batteries is justified. The paper is limited in scope to the study of two influential LP tests currently administered worldwide: TOEFL and IELTS. We will begin by pointing to the differences in British and North American traditions in language testing. This is followed by a brief review of the effect of the test methods on the measures of a construct. Then, the question of the research is discussed. In the method section, reviews of the tests concerned here will provide the basis for score comparisons across test batteries. Results of the comparisons will be reported and discussed in detail. Finally, the conclusion will sum up the discussions.

1.2 Test methods

There is a general belief that British and North American LP proficiency tests represent radically different approaches to language test development. North American tradition in language testing is heavily based on psychometric properties of tests. Issues such as reliability and concurrent and predictive validity are of particular interest in this tradition. Hence, objectivity of scoring and generalisibility of the results play a dominant role in the development of test methods. For example, multiple-choice items are often used in testing receptive skills to gain desired internal consistency, even if the test is expected to measure communicative competence as is the case in Functional Testing (Farhady 1980). Moreover, in order to achieve high inter rater reliability, the use of trained scorers and detailed specific instructions in conducting an interview are highly recommended for testing productive skills in this tradition.

When we examine the British tradition, it is observable that the emphasis is on the specification of test content and expert judgement. While reliability (degree of generalisibility of the results) receives less attention in this tradition, content and face validity are the major concerns of the test designers. It may follow that British tests enjoy more variability in their formats and include various communicative activities.

Different test methods might well affect the performance of the candidates taking the tests. The characteristics of test methods which influence test performance have long been studied by many researchers in language testing. Research has shown that test performance varies as a function both of an individual's language ability and of the characteristics of test methods. Some test takers, for example, might perform better in the context of a laboratory speaking to a microphone than they would in front of a panel of judges in an oral interview. Some test takers might find it easier to choose responses from among alternatives in a multiple-choice test of vocabulary than to complete an open-ended cloze format of a similar test. Completion of isolated sentences as opposed to completion of blanks in a text, live versus recorded speech, aural in contrast to written tests, are but a few examples of how methods of testing may vary. These characteristics of test methods may, in turn, influence the test performance, casting doubt on the reliability and validity of language tests. Controlling these characteristics thus becomes an important issue in the theory and

practice of language testing.

The study of test methods dates back to 1959 when Campbell and Fiske (1959) showed that method variance might influence the measures of a construct. They argued that a hypothetical large correlation between two traits, let us say A and B, and no correlation between traits A and C,might be a function of method variance common to the measures A and B and not to C, if the measures A and B are obtained by one method and that of C by another method. To control the method effect, they proposed a multitrait-multimethod (MTMM) design for validating tests. The main focus of the MTMM design is to separate trait and method factors. It recognizes that 'any test score is a function of both the trait it intends to measure and of the method by which it is measured' (Bachman and Palmer 1979:54). Therefore, the method involved in measuring might become as important as the trait it is intended to measure.

According to MTMM design, to observe the validity of a test, that is, to see whether the test is measuring what it purports to test, the application of more than one method seems necessary. If independent methods testing the same construct do tend to correlate highly, it is concluded that convergent validity is achieved. On the other hand, to achieve discriminant validity, i.e. to show that there are independent traits irrespective of the methods applied, introduction of more than one trait in the analysis is necessary. Low correlation between different traits indicates that they are really different from one another and hence discriminant validity is achieved.

As it stands, independence of methods is an important issue in validity as well as reliability studies. Convergence of independent methods claiming to test similar constructs is a proof of the validity of a test. However, in the case of reliability, convergence of similar methods is indicative of the reliability of the test. Since independence is a matter of degree, it may be concluded that reliability and validity can be considered to be on a continuum, depending on the degree of independence of test methods. That is,

> 'Reliability is the agreement between two efforts to measure the same trait through maximally similar methods. Validity is represented in the agreement between two attempts to measure the same trait through maximally different methods.' (Campbell and Fiske 1959: 83)

The MTMM design of Campbell and Fiske was influential for those interested to know whether the techniques testers use distort the results that they obtain. Bachman and Palmer (1981), for example, used a complex MTMM research design to investigate the comparative influences of two traits (speaking and reading) and three methods (interview, translation and self-rating). They found that scores from self-ratings loaded consistently more highly on method factors than on specific trait factors, and that translation and interview measures of reading loaded more heavily on method than on trait factors. Similar results were obtained in another study by the same researchers. Bachman and Palmer (1982) found that scores from both self-ratings and oral interviews consistently loaded more heavily on test method factors than on specific trait factors, while the scores from the multiple-choice and writing tests were least affected by method factors. A number of other studies have also examined the effect of test methods on test performance (see Alderson 1978, Bachman 1982, Lewkowicz 1983, Shohamy 1984, Chappelle and Abraham 1990).

What are the characteristics of test methods? The facets of test methods can be viewed from different perspectives. Bachman (1990:119) proposes a comprehensive framework for studying the facets of test methods. His framework comprises five main categories: facets of the testing environment, facets of the test rubric, facets of the input, facets of the expected response, and relationship between input and response. The large number of dimensions along which test methods vary in Bachman's framework are reflections of the variety of testing techniques that are used in language tests, and the ways in which these techniques vary.

Bachman's framework has been used for examining the various dimensions or facets of test methods in a large scale study, namely the Cambridge - TOEFL Comparability Study (Bachman, Davidson, and Foulkes 1993). This study offers an interesting suggestion: that different methods as diverse as Cambridge and ETS test batteries not only tap, to a large degree, similar abilities of the subjects in the sample concerned but also measure these abilities in much the same way. Among the findings of this study is the legitimacy of score comparisons across these two test batteries.

1.3 Scope of the present study

Bachman et al. (1993) suggested that score comparison across ETS tests and UCLES tests (CPE) could be made in a meaningful way. This would mean that institutional administrators across the Atlantic need not require separate test results for individuals who have already taken one of the test batteries. This will save time and money both for the individuals taking the tests and for the institutions offering the opportunities (admission, jobs etc.). If it is the case that score comparison is legitimate across Cambridge proficiency tests and ETS tests, the same comparison should also be possible between ETS tests (namely TOEFL) and IELTS (designed by UCLES). In addition, most universities in Australia and the UK require their non-native graduate candidates to provide a score on either TOEFL or IELTS as a proof of their proficiency in English. It seems that these institutions are practically equating the scores from TOEFL with those of IELTS.

In this research we are looking for the justification of score comparisons across TOEFL and IELTS. So the following questions are raised: Are TOEFL and IELTS comparable? Is there any consistent relationship between TOEFL and IELTS scores across time? Do preparation courses affect the performance of subjects in LP tests?

This study is also limited in scope to the study of Iranian graduate students' scores on TOEFL and IELTS between 1990 and 1992. Iranian graduate students who are intending to continue their studies by taking a PhD degree in English speaking countries are required to sit either TOEFL or IELTS. In many cases they sit both tests. The Ministry of Culture and Higher Education (MCHE) in Iran has developed a TOEFL-like test (MCHE) for screening the candidates before sitting the above tests. Only those who score above 50 (0-100 scale) on MCHE will be allowed to sit TOEFL or IELTS. The data presented here are based on the scores of those candidates who have sat all the three tests (IELTS, TOEFL, and MCHE) during 1990-1992.

2. Method

2.1 Reviews of proficiency tests

Prior to any discussion, analyses of the characteristics, activities and score bands as well as the underlying constructs of each test seem to be warranted.

2.1.1 TOEFL

The Test of English as a Foreign Language (TOEFL), a highly secure test, is the most widely administered, standardised , multiple-choice test of language proficiency (1963-1994). TOEFL is administered 12 times a year, a new equated form each month, at more than 1,100 centres in 170 countries and areas and its results are used by some 2500 universities and colleges in the US, Canada and other countries for a variety of academic subject areas. According to ETS (1992) some 1,178,193 students seeking admission to institutions in the United States or Canada took the test from July 1989 to June 1991. The test is designed to 'evaluate English LP of individuals whose native language is not English, most often those wishing to study in North American universities and colleges' (Stevenson 1987:79); it is recommended for students at 11th grade level or above.

2.1.1.1 The structure of TOEFL

The test comprises three sections (since 1976), each separately timed: Listening Comprehension (50 minutes), Structure and Written Expression (25 minutes), and Reading Comprehension and Vocabulary (60 minutes). All the items are in 4-MC format. TOEFL total scores range between 227-677 without any pass/fail scores. Nevertheless, institutions require different ranges of scores for different subject areas.

The TOEFL is, without a doubt, the most reliable as well as the most researched of all foreign LP tests, having been under constant revision and empirical research study for the past thirty years. The TOEFL Research series as of Summer 1993, consisted of 45 Research Reports and 6 Technical Reports. Over the years, TOEFL has been used as a criterion for the validation of other tests. Among the most recent attempts of this kind is the Cambridge - TOEFL Comparability Study (Bachman, et al forthcoming).

2.1.1.2 Reliability and Validity

The reliability of the test has repeatedly been reported to be satisfactory. Stevenson (1978) reports that ' the average reliabilities for 12 forms (administered in 1981-1982) are 0.89, 0.87, and 0.89 for the three sections, and 0.95 for the total score' (1987: 80). This is well within the desirable range for this type of test.

Validity of a test, by definition, depends on the extent to which a test measures what it purports to measure. TOEFL is intended to measure the English-language proficiency of non-native speakers of English who wish to study in North American universities. Hence, the content of the test should be representative of the social situations to which the examinees are expected to be exposed. The specification of such a context is not an easy task, given the wide range of TOEFL populations and target language-use situations. It seems that the traditional techniques of contrastive

analysis and error analysis are not appropriate for content selection of TOEFL. Like all proficiency measures, the content validity of TOEFL depends on the degree to which experts perceive it to be valid. Stevenson points out that:

> 'TOEFL does agree that content is best specified by experts, and does rotate membership in this group often to avoid stagnation or the dominance of one view, and leads to the reasonable conclusion, if not demonstration, that the content of TOEFL in general, is representative.' (1987: 81)

As for the construct validity of the test, we know that construct validity concerns 'the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs' (Bachman 1990: 255). The abilities involved in the construct of LP are theoretical, yet to be defined and agreed upon. Hence they constrain our efforts to test the extent to which we can make inferences about these hypothetical abilities on the basis of test performance. Unless we have a clear definition of the construct, we cannot claim to have measured it. TOEFL constructors seem to be very conservative in stating what construct they purport to test. For example, the TOEFL Bulletin of Information for TOEFL/TWE and TSE, 1992-1993 (ETS 1992: 3) states that the Vocabulary and Reading Comprehension section of the test 'measures ability to understand non-technical reading matter' in standard written English. It goes on to talk about the multiple-choice format of the questions implied, stated or otherwise. But it never explicitly defines the construct. As Peirce (1992:668) pinpoints, 'the construct of reading that is measured in the TOEFL reading test is not made explicit in the ETS literature'. Indeed ETS cannot make it explicit as there is no promising definition in the state of the art at present. Having said all this, there seems to be a general agreement in ETS that there exists a general proficiency factor which is divisible by skills and components.

2.1.2 IELTS

The International English Language Testing Service (IELTS) has been developed (1980-1994) jointly by the British Council and the University of Cambridge Local Examinations Syndicate (UCLES) to determine whether students' ability in English would meet the demands of a course of study in Britain and Australia. The early versions of the test (ELTS 1980-1989) comprised 6 subject specific areas in addition to a general section. The test reflects the ideas of communicative language teaching and is probably the first standardised communicative language test administered over a large population across the world. Some 37,455 non-native speakers of English are reported to have taken the test between 1981-1985 (Criper and Davies 1988). IELTS has been widely welcomed by the British and Australian universities as it claims to be a test of English for Specific Purposes (ESP).

Though the test was meant to be one of ESP, the final form includes an additional general section. The test follows the Munby (1978) communicative syllabus design. Carroll (1978) guided the test specifications on the basis of needs analysis. The analysis suggested a number of specific tests for different subject areas. However, in practice, large compromises and reductions were made, limiting the specific areas to six (from 1980 to 1989) and to three (since 1989), and perhaps only to one (from 1994), these changes being mainly determined by the British Council and UCLES, not by the students' needs.

2.1.2.1 The structure of IELTS

IELTS consists of two sections: General (G) and Modular (M). The general section consists of a listening test and an oral interview intended to test the oral skills. The Modular section, on the other hand, is intended to test the written skills: reading and writing. The modules are limited to two forms: Modules A, B, and C, for academic audiences, and Module GT, for non-academic general training purposes.

The listening part consists of thirty-five multiple-choice test items accompanied by a tape in four sections: 1) choosing from diagrams, 2) listening to an interview, 3) replying to questions, and 4) listening to a seminar. The interview is conducted face to face, individually, usually with a time lapse from the written test. It consists of two parts: general questions, and questions about candidates future plans. The subject is then assigned to one of the bands (1-9).

The overall formats of the modules (M1 = Reading) are the same. They all contain texts taken from books, journals, reports, etc., related to a specific subject area and involve testees in study skills necessary for academic studies, with the exception of the nonacademic area. There are all together 40 M-C test items in each module. The three academic modules are: Science and Technology, Life Science, and Arts and Social Sciences. Each student selects one module only. The Writing test has two questions in each module. The first question requires the testee to bring in his/her own experience and views on the basis of the reading texts. The second question is strictly limited to the information available in the text. Both tasks require the testees to write short paragraphs.

2.1.2.2 Reliability and Validity

There are no published statistics on IELTS except those reported by Alderson (1993) based on a trial test. Aside from the variations in the size of the trial population in different modules (not all students took every test in the battery), the reliabilities reported are acceptable. However, that of Module GT is questionable.

Table 1: Reliabilities Reported for IELTS: Trial Test: Alderson 1993

Tests	G1	MA	MB	MC	MGT	G2
Rel	0.86	0.90	0.91	0.88	0.79	0.87

G1 = Grammar Test; MA = Science and Technology Reading Test; MB = Life Science Reading Test; MC = Arts and Social Sciences Reading Test; MGT = Nonacademic Reading Test; G2 = Listening Test.

Alderson (1993) also reports the results of the reliabilities for the total test battery of listening, grammar, and reading tests ranging between 0.80-0.97, and that of the battery without the grammar test ranging between 0.76-0.96. Although the reliability of the total test battery declines in the absence of the grammar test, ' this decline is relatively unimportant, with the arguable exception of MGT, the General Training Model' (Alderson 1993: 215). The implication was that the grammar section should be dropped from the actual IELTS test. No reliability is reported for the total band

score.

A factor analysis of the test results reveals the emergence of a first dominant (general) factor followed by a second (writing) factor.

> 'In general, an analysis of reading, grammar, and listening yielded only one common factor. The addition of writing occasionally gave rise to a second factor.' (Ibid: 213)

Since Interview was not included in test analysis nor any other external criteria, it is difficult to predict what factors might have emerged had they been included in the analysis. The only statistics available in Alderson's (1993) report are the correlations between the two reading tests of the new (IELTS) test and the band score of the old ELTS subtests. The purpose of comparison 'was to enable the calculation of band scores to the test' (ibid: 214). There were significant variations in the relationship between the new and the old reading tests: readings MA correlated 0.39 while those of MC correlated 0.76.

The differences were justified on the assumption that the new IELTS test was an improvement on the old test and that the readings were not directly parallel to each other in content or topic.

Moderate correlations reported in the IELTS trial study between different modules support the ESP aspect of the test. IELTS does look and function like an ESP test. The test seems to be favoured more by its face validity than any other criteria. Due to the lack of published data, it is difficult to observe the extent to which the test measures what it purports to test. However, the factor analysis of the trial study does give evidence for the uni-factorial structure of the test.

IELTS seems to be based on a notion that proficiency is divisible by skill and as Alderson and Clapham (1992:164) report 'there are thus tests of the four macro-skills: reading, writing, listening, and speaking.'

2.1.3 MCHE

The Ministry of Culture and Higher Education Test of English Proficiency (MCHE) has been developed in Iran to assess the LP of Iranian graduate candidates who are awarded a scholarship to pursue their studies towards a PhD. At least three different versions of this test have been administered four times a year since 1989. The test comprises four multiple-choice sections: Listening Comprehension (30 items), Structure and Written Expression (30 items), Vocabulary (20 items), and Reading Comprehension (20 items). The total score is computed on the basis of the sum of the four sections (0-100). There is an additional writing (essay) section whose score (0-20) is reported separately. Due to administrative problems, the result of the latter section is not incorporated in this research.

There are no published data about the validity and reliability of this test. The structure of the test is very similar to that of TOEFL. The earliest version of MCHE was reported to have a correlation of 0.89 with TOEFL in 1989.

2.2 Subjects

The subjects were 1600 Iranian graduate students from different subject areas who sat for TOEFL and IELTS as well as for MCHE between 1990-1992. They were divided into two groups: Group A and Group B. Group A included students who sat for these tests between 1990 to early 1991 and for whom only the total scores for these tests were available. Group B included students who did the tests from early 1991 to mid 1992 and for whom both the total scores and the sub-section scores on each test were available. Only the scores of those who had done all the three tests were selected. Thus, only 113 and 103 subjects remained in Groups A and B respectively. Some students participated more than once in the tests. Only one score (the latest) of each student was counted for each test.

Moreover, most Group B subjects participated in TOEFL preparation courses during 1991-1992. Only a few participated in IELTS preparation courses. The IELTS sample materials, however, were distributed among all those from Group B who intended to sit for IELTS. The results reported here are based on 6 administrations of IELTS and 7 administrations of TOEFL.

3. Results

Relatively high correlations were found among Group A's scores on TOEFL and IELTS (table 2), while moderate correlations were found among Group B's scores on these tests.

Table 2. Correlations Between the Total Scores of the Tests: Group A Subjects

TESTS	TOEFL	IELTS
IELTS	0.8290	-----
MCHE	0.8339	0.7570

Table 3. Correlations Between the Total Scores of the Tests: Group B Subjects

TESTS	TOEFL	IELTS
IELTS	0.6671	-----
MCHE	0.6386	0.6072

By means of regression analyses score comparisons across tests were carried out. Tables 4 and 5 demonstrate the score comparisons across tests based on some of the key scores on MCHE.

Table 4. Score Comparisons Across Tests: Group A Subjects

TESTS	SCORES				
MCHE	50	60	70	80	90
IELTS	4.4	4.8	5.2	5.5	6
TOEFL	450	475	500	526	550

Table 5 Score Comparisons Across Tests Group B Subjects

TESTS	SCORES				
MCHE	50	60	70	80	90
IELTS	4 6	5	5 3	5 7	6
TOEFL	460	495	530	565	600

The rest of the results relate to Group B subjects Table 6 shows the mean score and standard deviation of the scores on each test

Table 6 Descriptive Statistics

	Mean Score	Standard Deviation
MCHE	52	10
IELTS	4 7	0 7
TOEFL	468	54

A full correlational matrix of the relationships between the different subsections of the tests is given in Appendix I A factor analysis was also conducted to find out the similarities between the two tests Table 7 shows the results of the factor analysis Varimax rotation extracted two factors All the subtests of IELTS and TOEFL loaded mainly on the first general factor associated with general listening ability The MCHE subtests loaded heavily on the second factor associated with general structure and reading comprehension

Table 7 Factor Analysis Rotated Factor Matrix

	FACTOR 1	FACTOR 2
MLC	47002	50915
MST	31146	63566
MVOC	25961	59368
MRC	10934	68662
IRC	68741	20749
IWR	62101	19941
ILC	69619	18214
ISP	49789	20632
TLC	75246	28392
TST	65528	49384
TRC	68478	44910

M = MCHE, I = IELTS, T = TOEFL, LC = Listening Comprehension, ST = Structure, VOC = Vocabulary, RC = Reading Comprehension, WR = Writing, SP = Speaking

Finally, to account for the effect of preparation courses (test effect), all the scores were converted to a scale of 0-20 so that the analysis of variance would become possible. A repeated-measures analysis of variance (MANOVA) was performed to find out whether there was any significant difference in the subjects' total score on the three different tests (TOEFL, IELTS, and MCHE). Table 8 illustrates the results of the MANOVA.

Table 8. Repeated-Measures Analysis of Variance

Source of Variation	SS	df	MS	F
Within Cells	355.82	204	1.74	36.95 *
Test	128.90	2	64.45	

* $p < 0.05$

The MANOVA detected a significant difference in the total test scores across the three batteries, suggesting the effect of the "test" factor. Of the three possible comparisons among the means, Tukey's WSD test shows that only the comparison between TOEFL and IELTS score was significant.

Table 9. Tukey Test of Differences Across Batteries

Test	TOEFL	MCHE	IELTS
	Mean = 10.73	Mean = 10.40	Mean = 9.22
TOEFL	--	0.33	1.51 *
MCHE	--	--	0.18
IELTS	--	--	--

* $p < 0.05$

4. Discussion

The reader should bear in mind that the intention of this research was not to carry out a full comparability study between IELTS and TOEFL. Rather, this research was conducted to show that these tests are not like apples and oranges and that score comparisons might be legitimate across these batteries. As far as face validity is concerned, the two tests might seem to be designed for different purposes: TOEFL as a general proficiency indicator and IELTS as an ESP one. Moreover, the researcher's personal interviews with a number of subjects (20) indicated that the majority of the testees preferred IELTS to TOEFL, believing that IELTS was a fairer indicator of their proficiency. The favourite section of IELTS, according to the subjects, was the reading section (ESP aspect), while the least favourite one was reckoned to be the listening part. The subjects, in general, thought that they had performed better at IELTS.

The question is whether the ESP colouring of IELTS makes it distinguishable from a

general proficiency test. Criper and Davies (1988) have shown that in spite of the intention of the designers of ELTS to create a multi-factorial structure test, the internal structure is in favour of a uni-factorial one. That is to say, general proficiency (whatever one may call it) is a better predictor of ELTS overall score. Alderson's (1993) trial study on IELTS also supports this idea. Although moderate correlations (0.51-0.67) between IELTS and TOEFL subtests reported in appendix I indicate that perhaps each test is testing something different- or rather, say in a different way- the factor analysis (table 7) indicates the dominance of a primary factor on which all the subtests of TOEFL and IELTS loaded and of which TOEFL listening comprehension loaded highest. This factor may well be interpreted as a general listening ability. The second factor, where MCHE's structure and reading comprehension loaded highest, could be interpreted as a general ability of reading comprehension and structure recognition. It may follow then that both TOEFL and IELTS acted unifactorially for the subjects concerned here. This is in accordance with previous research findings (Swinton and Powers 1980:15) that TOEFL acted unifactorially for less proficient groups. The TOEFL total mean score in this study is 468 which is far less than the average mean score for Farsi speakers (504) reported by ETS (1992).

The above discussions may lead us to the conclusion that IELTS and TOEFL share similar internal structure and may thus provide similar information of our testees' language ability. This allows us to do score comparisons across these tests in a rather meaningful way.

The results shown in table 4 are in accordance with most universities' expectations of the performance of non-native speakers on these two tests (see language proficiency requirement section of most UK and Australian Postgraduate Prospectus booklets). Score comparisons in table 4 indicate that a score of 6 on IELTS is equated with a score of 550 on TOEFL (the minimum requirement for allowing non-native speakers to enter into a non-linguistics department), while a score of 6.5 on IELTS is roughly equated with 600 on TOEFL (the minimum requirement for entering into a linguistics department). The comparisons in table 5, however, violate this equation. While changes in the less proficient subjects do not much affect the equation of the two scores, the changes in the scores for more proficient subjects (above 70 on MCHE scale) affect the equation in a meaningful way. Candidates who might have been accepted into a programme of study on their TOEFL score (600) would probably be rejected had their IELTS score (6) been taken into consideration. A closer comparison between Group A scores and Group B scores may suggest that subjects with approximately the same language ability performed differently in the two tests. Table 5 figures imply that subjects' (Group B) familiarity with the IELTS sample test had a slight improvement effect on the overall IELTS band score. They also imply that TOEFL preparation courses had a much higher improvement effect on the total TOEFL score. The effect is more striking for more proficient subjects.

Since score comparisons between Group A IELTS scores and Group B IELTS scores do not show much difference but the same comparisons between the two groups' TOEFL scores do show considerable difference, it may be concluded that TOEFL preparation courses had positive effect on the subjects' total TOEFL score. The overall MANOVA test shows that the effect of the factor "test" was significant. Moreover, the Tukey test suggests that the difference between the subjects' scores on TOEFL and IELTS was significant. It also implies that the subjects scored

significantly higher in TOEFL. This is in sharp contrast with what the subjects had earlier expressed in their interviews. Perhaps subjects' familiarity with the TOEFL format and their preparation courses were the main causes of this difference.

The Correlations reported in table 2 are not within one's expectation of the behaviour of similar LP tests. However, those reported in table 3 are well within one's expectation of the behaviour of LP tests. The difference might be due to the fact that scores reported here were gathered from different administrations of LP tests which might not have been equated to one another. So the difference might reflect the tests' unequated forms. It might also be due to the lower language ability of Group A subjects. Perhaps Group A subjects performed equally low at the two tests.

5. Conclusion

In this research we were looking for the justification of score comparisons across TOEFL and IELTS. We argued that since the internal structures of the two tests seem to be similar, tapping the same general proficiency factor, the tests may be comparable. It followed that score comparisons across the two test batteries are possible. The results of the comparisons suggested that although score comparisons across the two tests are possible, they might be affected by various factors across time. Factors such as test methods, subjects' familiarity with the test, LP preparation courses, and subjects' proficiency level might affect the score comparisons. This research was limited in scope to one native language only. Perhaps including the wide range of audience which these tests are addressing in the analysis would level the differences in score comparisons. Nevertheless, institutions using these test results should be cautioned about the relative comparability value of the test scores and should allow space for possible compromise of the band levels attached to the test scores. In short, score comparisons across LP tests are justified provided that possible extraneous factors affecting test scores are also taken into account.

Acknowledgements

I am grateful to Eric Glindinning for his comments on an earlier draft of this paper and to Dan Robertson for his help in doing the Tukey test. I alone am responsible for any mistakes.

References

Alderson C. 1978. 'A study of the cloze procedure with native and nonnative speakers of English'. Unpublished PhD dissertation. University of Edinburgh.

Alderson C. 1993. 'The relationship between grammar and reading in an English for academic purposes test battery' in D. Douglas and C. Chapelle (eds.) A New Decade of Language Testing Research. Alexandria, Virginia: TESOL: 203-219.

Bachman L. 1982. 'The trait structure of cloze test scores'. TESOL Quarterly 16,1: 61-70.

Bachman L. 1990. Fundamental Considerations in Language Testing. Oxford: OUP.

Bachman L. and A. Palmer. 1979. 'Convergent and discriminant validation of oral

language proficiency tests' in R. Silverstein (ed.) Proceedings of the Third International Conference on Frontiers in Language Proficiency and Dominance Testing. Carbondale, Ill.: Department of Linguistics, Southern Illinois University: 53-62.

Bachman L. and A. Palmer. 1981. 'The construct validation of the FSI oral interview'. Language Learning 31,1: 67-86.

Bachman L. and A. Palmer. 1982. 'The construct validation of some components of communicative proficiency'. TESOL Quarterly 16,4: 449-65.

Bachman L., F. Davidson and J. Foulks. 1993. 'A comparison of the abilities measured by the Cambridge and Educational Testing Service EFL test batteries' in D. Douglas and C. Chapelle (eds.) A New Decade of Language Testing Research. Alexandria, Virginia: TESOL: 25-46.

Bachman L., F. Davidson, K. Ryan, and I.C. Choie. Forthcoming. The Cambridge-TOEFL Comparability Study. Final Report. Cambridge: UCLES.

Campbell D. and D. Fiske.1959. 'Convergent and discriminant validation by the multitrait-multimethod matrix.' Psychological Bulletin 56: 81-105.

Carroll J. 1978. English Language Testing Service. Specifications. London: The British Council.

Chapelle C. and R. Abraham. 1990. 'Cloze method: what difference does it make?' Language Testing 7,2: 121-146.

Criper C. and A. Davies. 1988. ELTS Validation Project Report 1. Hertford: The British Council and UCLES.

ETS. 1992. Bulletin of Information for TOEFL/TWE and TSE, 1992-93. Princeton, NJ.: Author.

Farhady H. 1980. 'Justification, development, and validation of functional language testing'. Unpublished PhD dissertation. University of California Los Angeles.

Lewkowics J. 1983. 'Method Effect in Testing Reading Comprehension: a Case of Three Methods'. Unpublished MA dissertation. University of Lancaster.

Munby J. 1978. Communicative Syllabus Design. Cambridge: CUP.

Peirce B. 1992. 'Demystifying the TOEFL reading test'. TESOL Quarterly 26/4: 665-691.

Pike L. 1979. An Evaluation of Alternative Item Formats for Testing English as a Foreign Language. TOEFL research report 2. Princeton, NJ.: ETS.

Shohamy E. 1984. 'Does the testing method make a difference? The case of reading comprehension'. Language Testing 1,2: 147-70.

Stevenson D. 1987. 'Test of English as a Foreign Language' in C. Alderson, K. Krahnke, and C. Stanfield (eds.) 1987:79-81. Reviews of English Language

Proficiency Tests. Washington D.C.: TESOL.

Swinton S. and D. Powers. 1980. Factor Analysis of the TOEFL for Several Language Groups. Princeton, NJ.: ETS.

Appendix 1
Correlational Matrix Among Subsections of the Three Tests

	MLC	MST	MVOC	MRC	IRC	IWR	ILC	ISP	TLC	TST	TRC
MLC	1.0000										
MST	.4884**	1.0000									
MVOC	.4663**	.4055**	1.0000								
MRC	.4277**	.4709**	.4486**	1.0000							
IRC	.4584**	.3333**	.3278**	.1971	1.0000						
IWR	.4178**	.2991*	.2445*	.2524*	.5267**	1.0000					
ILC	.4678**	.3245**	.2477*	.2519*	.5473**	.4675**	1.0000				
ISP	.3508**	.2307*	.2869*	.2446*	.2897*	.4256**	.4823**	1.0000			
TLC	.5445**	.4181**	.3637**	.2076	.5211**	.4637**	.5619**	.4408**	1.0000		
TST	.4710**	.5593**	.4392**	.4064**	.5329**	.5069**	.4898**	.3944**	.6870**	1.0000	
TRC	.4468**	.5433**	.4827**	.3314**	.6072**	.4659**	.5044**	.3816**	.6877**	.7637**	1.00

N of cases: 103 1-tailed Signif: * - .01 ** - .001

M = MCHE, I = IELTS, T = TOEFL, LC = Listening Comprehension, ST = Structure, VOC = Vocabulary, RC = Reading Comprehension, WR = Writing, SP = Speaking

BEST COPY AVAILABLE

THE UNIVERSITY OF EDINBURGH TEST OF ENGLISH AT MATRICULATION: VALIDATION REPORT

Tony Lynch (IALS)

Abstract

This paper reports on a two-part evaluation of the Test of English at Matriculation (TEAM) in use at the University of Edinburgh. Separate samples of candidates' scores were used to assess (1) TEAM's concurrent validity with other measures of English language proficiency and (2) its predictive validity in relation to academic outcome. These statistical comparisons established strong correlations with existing tests, particularly the English Proficiency Test Battery, and suggest that TEAM performs predictively as well as other measures, scores on the TEAM listening subtest being especially indicative.

1. Background

Since the early 1970s the University of Edinburgh's policy has been to provide in-session English tuition for non-native students who have fulfilled the linguistic entry requirement but are thought likely to gain, in terms of improved course performance, from further language support. The entry requirements vary among the faculties at Edinburgh, but most currently take IELTS 6.0, TOEFL 550 or English Proficiency Test Battery (EPTB, Version D) 40.0 as the minimum for acceptance.

TEAM is the most recent of three matriculation tests that have been used by the University at matriculation to identify students who are likely to be at risk linguistically and who should receive English language support. The first was the English Language Battery (ELBA), which was used until 1982; the second was the British Council/UCLES ELTS test, taken at matriculation in the period 1982-86, while the ELTS Validation Project was under way at the University of Edinburgh. As the project approached its end, a decision was taken by the University's English Language Testing and Tuition committee to replace ELBA (a multiple-choice test of grammar, vocabulary and reading) with a test that would also sample students' listening and writing.

TEAM was introduced for the academic session 1987-88 and piloted over two years in tandem with ELBA. It consists of four parts: a vocabulary test, a dictation test, a reading comprehension test and a writing test. In deciding whether or not to refer students for the in-session courses, their overall average score is interpreted as follows: less than 50% - at least 50 hours' tuition required; 50-59% - tuition strongly recommended; 60% and above - tuition may be recommended, depending on subtest scores. In comparing TEAM with ELBA it was therefore of particular importance to

compare the distribution pattern among the score bands used as the basis for referral (see Table 1).

Table 1 Student distribution (%) by score band: ELBA and TEAM 1987-89

ave	ELBA	TEAM
<50	40	29
50-59	24	29
60-69	16	19
<69	20	23
	100	100

The key score bands, i.e. those interpreted as indicating that English in-session tuition is 'required' and 'strongly recommended', show a broadly similar distribution of students on the two tests (64% on ELBA and 58% on TEAM). Concurrent performances on the two tests by matriculating students during the two-year trial (n=95) showed a Spearman correlation of .81 ($p < 0.01$). The pilot study report (IALS 1989) concluded that TEAM was in general terms an adequate replacement for ELBA, yielding a similar picture of students' English proficiency.

TEAM has been in independent use as the University's matriculation test of English since the 1989/90 academic session. When advising students and staff of results, we may be asked about the relationship between TEAM and other measures, particularly the test that students have taken in their home country, and about how TEAM scores relate to academic success. A two-part study was therefore undertaken to investigate these two issues - TEAM's concurrent and predictive validity.

2. Concurrent validity

2.1 Method

Data for the study of concurrent validity was available in IALS archives in the form of the test scores of students attending our pre-sessional EAP courses over the period 1982-92 who had been required to take a test at the end of the pre-sessional for acceptance onto their subject courses (n=358). These records allowed comparison of individuals' performances on at least two tests: ELTS or EPTB (taken in Scotland to achieve acceptance onto the subject course), and ELBA or TEAM (taken at matriculation). In addition, approximately a quarter of the sample (n=80) had taken an IALS cloze reading test for EAP placement purposes.

Although all these tests were taken in September of the relevant year, it should be emphasised that this first part of our validation project cannot claim to assess strict concurrent validity, since the test data it investigated was not gathered under controlled conditions. With the exception of a cohort of students who were included in a three-way comparative study of ELTS/ELBA/EPTB for the ELTS Validation Project in 1982, the test candidates in the IALS pre-sessional sample did not take their tests on the same day. The interval between test sessions ranged from one to two weeks in the case of the EPTB, ELBA, TEAM and ELTS, and up to three weeks in the case of the Cloze test. However, as TEAM scores are interpreted in an approximate way (firstly as the individual student's average over the four subtests,

and secondly through the use of decile score bands) it was considered reasonable to aim for a broad-brush comparison with other tests. Table 2 shows the breakdown of the pre-sessional sample into inter-test comparisons.

Table 2. Inter-test comparisons in the pre-sessional sample 1982-92 (n=358)

ELTS x EPTB x ELBA	24
ELTS x EPTB	45
EPTB x TEAM	194
ELTS x TEAM	36
ELTS x Cloze	30
ELBA x Cloze	26
TEAM x Cloze	80

It will be noted that comparison figures exceed the subject total of 358, since a number of students took more than three tests. Although this pre-sessional sample contained no direct comparison of ELBA and TEAM, figures were available on students (n=95) taking both tests concurrently at matriculation in 1987 and 1988 for the TEAM pilot study (IALS 1989).

2.2 Results and discussion

Table 3. Means, standard deviations, minimum and maximum scores (1982-92)

	ELTS	EPTB	ELBA	Cloze	TEAM
mean	5.78	39.12	51.70	66.82	50.69
s.d.	0.75	7.56	14.41	20.22	11.85
min.	3.50	23.00	18.00	8.00	25.00
max.	7.00	59.00	84.00	120.00	86.00
poss.	9.00	65.00	100.00	147.00	100.00

These mean scores indicate broad similarity with the standard interpretation scale in use at British universities to compare EPTB with ELTS for acceptance on a university course, in which ELTS 6.0 is regarded as equivalent to EPTB 40.0 (and TOEFL 550). It also confirms that, taken over the five academic sessions since its initial trialling in 1987, TEAM has achieved reasonable similarity with its predecessor, ELBA.

Table 4. Pearson correlation matrix for the five tests

	ELTS	ELBA	Cloze	TEAM
EPTB	.74	.83	.84	.94
ELTS		.72	.70	.72
ELBA			.93	.81*
Cloze				.77

($p < 0.01$ in all cases)
*source: IALS (1989)

A number of points may be made about the correlation values shown in Table 4. Firstly, although we have already drawn attention to the restricted sample size in

some inter-test comparisons, even the smallest subsample (n=24), for ELBA and EPTB shows a correlation (.83) very close to the .85 reported for a much larger sample (n=430) in Criper and Davies (1988). So these IALS pre-sessional students may be regarded as typical of the wider population of international students entering universities in Britain.

Secondly, the test that achieved the lowest correlation *vis-a-vis* the other four tests was ELTS, with figures ranging from .70 with Cloze to .74 with EPTB. One possible reason is that ELTS is the only test of the five to examine oral proficiency, through interview; it may be that performance on speaking varies among candidates in ways not reflected by their patterns of scores on the other ELTS subtests. This would in fact be the converse of the case of the two pairs of tests in Table 4 that are most similar in focus, if not format: TEAM and EPTB (testing listening, reading and writing) and Cloze and ELBA (testing grammar, vocabulary and reading); these pairs have high correlations - .94 for EPTB/TEAM, and .93 for ELBA/Cloze. Further possible weakening influences on correlations with ELTS are the low reliability of the interview module, commented on in the ELTS Validation Report (Criper and Davies 1988), and potential inconsistencies between performances on the original five-module ELTS and the revised four-module IELTS, introduced in 1989.

Cross-tabulation of scores allows us to confirm the existing interpretation scale of EPTB and ELTS, and to extend it to include TEAM and the Cloze, as shown in Table 5.

Table 5. Comparison across test score bands

TEAM	ELTS	EPTB	Cloze
80%	7.5	55.0	110
70%	7.0	50.0	100
60%	6.5	44.0	85
50%	6.0	40.0	70
40%	5.5	38.0	60
30%	5.0	36.0	50
20%	4.5	34.0	40

Two caveats are in order here, since there is a risk that the score interpretation in Table 5 will be seen as in some sense the 'principal result' of this investigation of concurrent validity. Firstly, we have already emphasised the restricted sample size available for some inter-test comparisons, even though we know that results from the smallest do bear comparison with those of the larger ELTS Validation Project sample. Secondly, the reader / user of the interpretative table should bear in mind when converting one test into another that, with the exception of the Cloze, the result of all the tests in this study takes the form of an overall score combining marks on a number of subtests; this inevitably conceals what may be markedly different patterns of achievement on the subtests, which need to be taken into account in assessing a student's ability to carry out the various academic tasks that postgraduate courses demand.

However, since the purpose of TEAM is diagnostic, to evaluate likely need for in-session language support, and not to act as a pass/fail criterion for acceptance onto a course, these results suggest that TEAM stands up well to detailed comparison with

other measures of international students' English. In particular, its high correlation with EPTB (.94) indicates a firm basis for direct comparison of performances on those two measures.

3. Predictive validity

3.1 Establishing criteria

Having discussed the extent to which TEAM scores reflect achievement on other language tests, we now turn to the issue of predictive validity. In so doing, we seek an answer to the other question we are sometimes asked by academic staff, which might be paraphrased as 'What do TEAM scores tell us about how well this student will do on our course ?' Before considering the details of this second part of our study, it is worth briefly reviewing some of the main problems in establishing predictive validity.

The first is the question of what criterion to select as a basis for measuring academic success. One might make a simple two-way distinction of Pass or Fail, but this would blur the gradations of academic performance that are an established part of the British system of percentage marking. It would also inevitably conceal differences between the student who achieves Distinction and one who scrapes a borderline pass.

More specifically, where a postgraduate course has three possible outcomes, as is the case with most courses at Edinburgh, of Pass at Master's level, Pass at Diploma level and Fail, there arises the issue of how to categorise the Diploma Pass. Should we regard it as a form of failure and take the Master's Pass as the only real success? Or should one accept the arguments of the departmental staff who regard a Diploma Pass on their course as a mark of solid achievement and a Master's Pass as a bonus? Our experience is that staff attitudes to the status of the Diploma Pass varies among (and also within) departments.

Thirdly, any comparison of language test scores with outcomes in a range of academic fields involves the assumption that all the departments in an institution are working to the same academic standards. Our purpose here is to assess the predictive validity of an English language test, rather than to attempt an academic audit, and we will therefore assume that a Diploma pass in one academic subject is the same as one in another. If this is a fiction, it seems to us a necessary one.

3.2 Method

The data for analysis comprised the TEAM scores of students matriculating at Edinburgh in the three sessions 1989-90, 1990-91 and 1991-92 for one-year taught postgraduate courses, primarily Diploma/M.Sc courses (n=291). There were two main reasons for our decision to focus on these students, rather than on those beginning research degrees. The first was related to the diagnostic aim of TEAM; the University of Edinburgh has always assumed that students on 12-month courses run a greater risk of failure than those taking research degrees, which involve a different and perhaps less intensive pattern of study, and certainly a longer period in which to remedy any language weaknesses. The second reason was a practical one: at the time of our study, data on Diploma./M.Sc. outcome was available for the three annual

intakes after 1989, whereas very few of the research students first matriculating in 1989 would have had time to complete their research.

In order to gather data on outcome, a questionnaire was sent to Faculty officers dealing with postgraduate students. The form comprised a simple checklist for each academic session, listing TEAM candidates from the relevant Faculty; staff were asked to indicate one of four outcomes - Master's Pass, Diploma Pass, Fail, or research; a final column provided space for 'other comments'. Table 6 summarises the distribution among the three taught-course outcomes.

Table 6. Overall M.Sc. success / failure rates of TEAM candidates 1989-92

M.Sc. pass	Diploma pass	failure	TOTAL
230 (79%)	34 (12%)	27 (9%)	291

The 9% failure rate may appear high and it is important to make clear precisely what we have included under that heading. The Faculty responses to our questionnaire provided a variety of comments on non-completion as opposed to a Fail: e.g. 'withdrew before resits', 'returned home because of family problems', 'discontinued', 'withdrawn during study'. We are also aware of cases where students started an M.Sc. course but experienced such difficulties with English that they left the University after the first few weeks of the Autumn Term; officially there was 'no record' of their participation in the course.

Failure is a sensitive issue in any area of life and there are obvious pressures on departments not to fail students: technically, a student who withdraws (or is withdrawn) from a course has not 'failed', but withdrawal can be taken as an indication that an individual would have failed. As Criper and Davies (1988) point out, even when medical or family reasons for non-completion are cited, it may well be in order to save embarrassment, either personal or institutional. Given the inevitable uncertainties of explicit and implicit failure and the possible hidden influence of language problems on non-completion, we decided to adopt a broad definition of 'failure' in this study, and to include in that category both outright Fails and non-completions. Although there might be objections that this has exaggerated the failure rate, it is clear from Table 7 that our categorisation has in fact resulted in an overall distribution almost identical with that found in the ELTS Validation Report.

Table 7
Overall success / failure rates on Master's courses:
ELTS validation sample (n=502)

M.Sc. pass	Diploma pass	failure
81%	12%	7%

We can assume, then, that the decision to combine 'Fail' and 'non-completion' has not skewed the pattern relative to ELTS; this will allow us to compare the predictive validity of the two tests with some confidence.

3.3 Results and discussion

Table 8. TEAM: means, standard deviations, minimum and maximum scores Master's course sample 1989-92

	Vocab.	Dict.	Reading	Writing	Ave.
mean	53.38	63.81	51.68	63.76	59.62
s.d.	14.31	21.26	25.94	16.08	15.03
min.	6.00	9.00	0.00	15.00	14.00
max.	100.00	100.00	100.00	100.00	99.00

The overall TEAM average score is higher than the 50.69 figure in the concurrent validity sample (Table 3), but this can be explained by the differences between the two populations: the students whose scores are presented in Table 3 had been required to attend pre-sessional tuition and also included research students, while the figures in Table 8 are those of Master's course students attending the matriculation test of English, the majority of whom were not required to take tuition prior to subject course entry. So one would expect the students in the matriculation sample to produce higher scores overall.

When the overall average TEAM scores are banded by deciles and compared with outcome (Table 9), we find some initial evidence of a relationship between language proficiency as measured by the matriculation test and success on the departmental course.

Table 9. Distributions of TEAM Average scores and academic outcome Master's course sample 1989-92

TEAM Ave.	Master's pass		Diploma pass		failure		Total
< 30%	1	(33%)	-		2	(67%)	3
30-39%	8	(50%)	3	(19%)	5	(31%)	16
40-49%	32	(68%)	9	(19%)	6	(13%)	47
50-59%	58	(75%)	11	(15%)	8	(10%)	77
60-69%	55	(81%)	8	(12%)	5	(7%)	68
70% or more	76	(95%)	3	(4%)	1	(1%)	80
overall	230	(79%)	34	(12%)	27	(9%)	291

The failure rate decreases with increasing English proficiency, falling from 67% at TEAM scores below 30% to a mere 1% of failure at TEAM scores of 70% or more. Conversely, Master's pass rates rise from 33% for those scoring below 30% on TEAM to 95% for those achieving above 69% on TEAM. The watershed of better-than-average chances of passing at Master's or Diploma level is around TEAM 60%. In considering the general pattern of the relationship between TEAM results and success or failure on the subject course, we might also look at the test/outcome findings of the ELTS Validation Study (Table 10).

Table 10. Distributions of overall ELTS scores and academic outcome: ELTS project sample (n=720)

overall band	failure
up to 4.0	57%
4.5	33%
5.0	33%
5.5	30%
6.0	19%
6.5	6%
7.0	5%
mean failure rate	22%

It is important to note that in Table 10, the apparently very high 'failure' rate was based on a definition of failure that encompassed both Fails and Diploma passes, and so in order to compare these findings with those of our own predictive study, we have to combine the relevant means in Table 9 - 12% Diploma passes and 9% failures, giving 21%. So again there is a close similarity between the ELTS findings and those for TEAM. Criper and Davies (1988: 92) concluded that ELTS 6.0 could be regarded as 'the dividing line between an acceptable and unacceptable risk of failure'. For our Master's course sample it appears that the cross-over point is in the 50-59% TEAM band and that this applies both to the chances of getting a pass at Diploma level and also to the likelihood of failure (whether outright Fail or non-completion). The evidence is, then, that the level of English proficiency below which a student stands an above-average chance of not passing the degree for which they are registered is 6.0 on ELTS and 50-59% on TEAM.

Overall, then, the evidence of Tables 9 and 10 is that the pattern of performance in the Edinburgh TEAM sample was similar to that in the larger ELTS sample: one in five non-native students ran a risk of not getting their Master's degree.

Having discussed the global pattern of TEAM average scores, we now consider performance on the four TEAM subtests. The figures in Table 11 suggest that some parts of TEAM perform better than others as predictors of outcome.

Table 11. 'Failures' by TEAM subtest bands (all figures %)

	Vocab	Dict	Read	Wri
<30	50	50	15	25
30-39	14	16	13	0
40-49	10	20	14	13
50-59	10	6	5	13
60-69	5	4	3	13
70 or more	4	5	5	4

The vocabulary test and the dictation test both produce clines of increasing scores and falling rates of failure. However, the rather flat spread of scores on the reading subtest means that it does not discriminate sufficiently at lower levels; the chances of failure are not differentiated among reading scores up to 50%. On the other hand, the 50-59% band does appear to mark a division, with a decline in failure rates with

TEAM scores above 50%. The writing test produces a level bunching of students who performed relatively well on that subtest (40-69%) but nevertheless failed or did not complete their degrees.

Table 12. Mean TEAM subtest scores (%) by outcome

	Master's pass	Diploma pass	failure
Vocabulary	57.03	48.07	47.70
Dictation	67.80	57.65	47.70
Reading	54.01	37.71	40.44
Writing	67.54	61.47	58.52
Ave	63.38	53.29	49.59

On the evidence of the results in Table 12, the dictation subtest produces the clearest differentiation among the three outcomes, with a mean interval of some 10%. The vocabulary section of TEAM appears not to discriminate sufficiently between Diploma Pass and failure. Scores on reading are erratic and those on the writing subtest have a restricted range.

Table 13. Pearson correlations: TEAM subtest scores with outcome

Vocabulary	0.24
Dictation	0.31
Reading	0.22
Writing	0.19
Average	0.32

($p < 0.01$ in all cases)

Dictation emerges as the subtest with the closest association with students' eventual success on their course, and the correlation of .32 for the association between Average and outcome is comparable with those reported in the ELTS Validation Report of .34 between outcome and ELTS taken at home, and .35 between outcome and ELTS retaken in Britain. The extent to which each of the subtests can be said to have contributed to eventual success is shown in Table 14. The dictation score is the only statistically significant coefficient.

Table 14. Regression analysis - logistic estimates (depend. variable: 1 = M.Sc./Dip. Pass; 0 = failure)

	coefficient (*t* tests)
Vocabulary	0088
	(.866)
Dictation	0223
	(3.259)*
Reading	-.0012
	(-.224)
Writing	0029
	(.358)

* significant at the 1% level

The fact that the dictation subtest performs best as a predictor is of particular interest. One might have expected that, since the assessment of performance on postgraduate courses is based predominantly on written assignments (essays, projects, examination and dissertation), it would be measures of text skills (reading and/or writing) that reflected subject course performance better than a test of listening comprehension. Foreign language use being complex rather than simple, it seems likely that the link between listening and outcome is an indirect one. It is evident to subject staff and language tutors alike (and to the students themselves) that individuals who, from the very beginning of the first term of a one-year taught course, have difficulty in understanding lecturers are likely to fall behind in their grasp of the factual and conceptual content of the course and may never catch up in what is a relatively short and intensive period of study.

From the wider perspective of research into second language acquisition (e.g. Faerch and Kasper 1986; Rost 1990), listening is regarded as a powerful source of input to the acquisition process, provided that the messages are comprehensible. But second language users who are unable to cope with the pace and complexity of lectures may experience a multiplier effect - losing confidence in their ability to understand spoken English, therefore becoming more anxious about lecture comprehension and note-taking and all the while appearing to lose ground to their peers who are able to follow the language and content of the lectures. More generally, the comprehension barrier can cut them off from the host culture, and this may in turn contribute to the loneliness and homesickness that can later surface as 'family' and 'medical' reasons for withdrawal from the course. Interestingly, there is North American evidence that aural comprehension ability exerts a strong influence on academic success even in the first language; Oxford (1993) cites an extensive survey by Conaway (1982), which found that poor listening comprehension was a more significant factor in academic failure than poor reading comprehension and low academic aptitude.

Our analysis of the TEAM scores suggests that, as in the L1 case, listening skills tapped by the dictation subtest may be a key element in academic success for international postgraduates on one-year courses. However, it could be that what enables students to respond well to the specific demands of a dictation is a more

general language proficiency factor and not only aural comprehension; proponents of dictation such as Oller (1976, 1979) have long argued that a dictation test is an effective probe of the learner's expectancy grammar, providing insight into general language competence.

The measurable predictive power of TEAM overall, like that of other British language tests, is relatively limited. Criper and Davies (1988) established a correlation of approximately .3 between overall ELTS scores and academic outcome, and described that as typical of similar investigations of predictive validity. It is true that a number of North American studies (reviewed in Graham 1987) have reported correlations as high as .5 between English proficiency scores (usually TOEFL) and academic performance, but the measure of the latter has tended to be the student's first-semester grade-point average (GPA), rather than performance later in their course career. It may also be significant that the US studies have tended to focus on undergraduates rather than graduates, since the demands placed on non-native users by the two types of degree are likely to be different.

However, to conclude that TEAM accounts for some 10% of the variance in academic performance across the sample as a whole does not exclude the possibility that (in)ability in English may represent much more than 10% of the difficulty that linguistically weak students encounter in following their degree course. 'It is feasible that the low correlations between language level and final outcomes mask a non-linear relationship: that the effect of language increases steeply at lower levels' (Criper and Davies 1988: 91).

4. Conclusions

On both issues investigated in this study, concurrent and predictive validity, TEAM bears comparison with established and more widely used tests. We have found reasonable grounds for confidence in the interpretation of TEAM scores in terms of its concurrent validity relative to other measures of academic English proficiency, particularly EPTB. Since EPTB is offered as an alternative to IELTS to pre-sessional students studying in Edinburgh for acceptance onto a university course, the evidence of a close relationship between EPTB and TEAM is an especially valuable finding of this study of concurrent validity.

As a predictive instrument, TEAM performs on a par with the original version of ELTS, achieving a correlation of .32 between overall TEAM average score and academic outcome. We have stressed that this is an association across the whole population, encompassing a wide range of ability; a reasonable case can be made that for students with relatively weak English - those likely to be identified as requiring language tuition - the influence of language ability (and listening in particular) will in fact have a substantially greater influence on their particular performance on a course than is apparent from the 10% global figure.

In both the validation studies reported here, we have compared TEAM's performance with the original version of ELTS. We await with interest the publication of the ongoing UCLES validation study of IELTS (Ferguson and White, in progress), which will allow us to relate TEAM more closely with the current version of the test.

Although TEAM appears to do as well as other tests, there is a need to revise some of its subtests; while the TEAM dictation score acts as a reasonable predictor of academic outcome, our analysis has demonstrated that the reading and writing subtests require adjustment in order to raise their predictive power. A revised version of TEAM has now been introduced and we intend to evaluate the effects of those revisions in a future study.

Acknowledgment

I would like to express my thanks to the staff in the Registry and Faculty Offices of the University of Edinburgh for their assistance with the data collection for the predictive validation component of this research, and to two EWPAL readers for their helpful comments on an earlier version of this paper.

References

Conaway M. 1982. 'Listening: learning tool and retention agent'. In A. Algier and K. Algier (eds.) Improving Reading and Study Skills. San Francisco: Jossey-Bass.

Criper C. and A. Davies. 1988. Edinburgh ELTS Validation Project Report. British Council / University of Edinburgh.

Faerch C. and G. Kasper. 1986. 'The role of comprehension in second language learning'. Applied Linguistics 7/3: 257-74.

Ferguson G. and E. White. (in progress) A Predictive Validity Study of IELTS. University of Cambridge Local Examinations Syndicate / University of Edinburgh.

Graham J. 1987. 'English language proficiency and the prediction of academic success'. TESOL Quarterly 21/3: 505-21.

IALS. 1989. English Language Testing and Tuition Report: TEAM Pilot Study. Mimeo. University of Edinburgh.

Oller J. 1976. 'Evidence for a general proficiency factor: an expectancy grammar'. Die Neueren Sprachen 2: 165-174.

Oller J. 1979. Language Tests at School: a Pragmatic Approach. Longman: London.

Oxford R. 1993. 'Research update on teaching L2 listening'. System 21/2: 205-11.

Rost M. 1990. Listening in Language Learning. Longman: London.

SOME ASPECTS OF 'FOREIGNNESS' IN THE PRONUNCIATION OF UPPER INTERMEDIATE ENGLISH STUDENTS OF SPANISH

Carmen Santos Maldonado (DAL)

Abstract

The present study was designed to address the following three questions related to the pronunciation of Spanish as a foreign language: 1. Is the pronunciation of upper intermediate learners 'fossilized'? 2. Is there a relationship between 'quality' in pronunciation and 'amount of improvement'? 3. When judging degrees of 'foreignness', are linguistically trained native judges 'harsher' than linguistically naive native judges? Some upper intermediate learners of Spanish were recorded 'before' and 'after' a programme in Spanish pronunciation. Then the same ten 'before' and 'after' pairs of sentences of each student were carefully randomized and rated for quality of pronunciation by native speakers of Spanish. Results suggest, on the one hand, that phonological fossilization is present but does not affect everybody to the same extent; on the other hand, that even at this high level some students can benefit considerably from pronunciation training. We conclude by discussing the potential value of reconsidering the place of pronunciation in language teaching in a university setting.

1. Background

There can be few people learning a foreign language who do not wish to become proficient in all areas of the language: grammar, vocabulary and phonology. Although learners make mistakes at all three levels, there is a belief that grammar and vocabulary, but not phonology, can always improve. This view is widely reflected in foreign language teaching practice, and, while grammar and vocabulary are extensively worked on in the classroom, pronunciation is largely neglected soon after the initial stages of learning.

Although research in Interlanguage Phonology (IP) has been sparse, it is possible, according to Tarone (1987), to trace two major issues in this field. The first one refers to the nature of the processes shaping IP, processes such as transfer - both positive and negative - (Brière 1966; Altenberg and Vago 1987), first language acquisition factors (Wode 1976; Hecht and Mulford 1987), overgeneralization (Tarone 1987), approximation or phonological translation (Flege 1980, 1981, 1987a) and avoidance (Tarone 1987). The other major issue is the phenomenón of fossilization.

Attention to the concept of fossilization was drawn mainly by Selinker (1972) in relation to the Interlanguage Hypothesis. He defines fossilizable linguistic phenomena as those which are not likely to improve any further, 'no matter the

amount of explanation or instruction the learner receives in the TL' (Selinker 1972: 215). Fossilization is therefore the cessation of acquisition of any further knowledge before the learner has acquired a native-like level of performance. In the domain of syntax, fossilization results in ungrammatical sentences, and in the domain of phonology it results in a 'foreign accent'. As adult L2 learners know only too well, it is virtually impossible to be taken for a native speaker in a conversation in the target language; sentences like 'you've got a very good accent', however reinforcing they may sound, do nothing but underline the fact that the non-nativeness has been noticed. This everyday experience is corroborated by Scovel (1969), who states very firmly that no adult ever achieves perfect native pronunciation in an L2. Asher and García (1969) go even further by noting that even many children coming into contact with the target language as late as the age of six do not achieve completely native pronunciation. Not a very encouraging outlook! Without any doubt age is a primary factor influencing pronunciation, but it is not within the scope of this paper to examine the constraints on pronunciation mastery fully. An adequate discussion of the influence of these constraints has to bear in mind factors such as: age and the 'critical period' issue (Krashen 1973; Snow and Hoefnagel-Hohle 1982; Loewenthal and Bull 1984; Flege 1897b), second language input (Krashen 1985; Tahta, Wood and Loewenthal 1981b), motivation (Gardner and Lambert 1972; Suter 197o; Purcell and Suter 1980), affective factors (Guiora, Beit-Hallahmi, Brannon, Dull and Scovel 1972; Schumann 1976; Harder 1980), aptitude (Flege 1981, 1987b), sex (Tahta, Wood and Loewenthal 1981a, 1981b) and the first language (Suter 1976; Purcell and Suter 1980).

Not all researchers, however, put forward a pessimistic view about the possibility of acquiring good L2 pronunciation. Neufeld (1977) obtained very positive results with English-speaking young adults who were tested for their ability to reproduce, with the accuracy of a native speaker, the phonological features of two non-Indo-European languages: nine out of twenty L2 learners convinced three native speakers that the target language was their native language. Although Neufeld himself acknowledges the very restricted circumstances under which the experiment took place he is ready to admit that adults can sometimes learn an L2 pronunciation system with the accuracy of a native speaker, an idea that has very interesting implications for the teaching of pronunciation.

We would like to make two more points about fossilization in learning pronunciation. The first one refers to the definition of fossilization proposed by Selinker (1972), that an item becomes fossilized when it does not improve with instruction. It seems to us that this definition may run into a vicious circle because how do we know that no further instruction is required? How do we know that no further improvement will take place? Indeed if instruction stops maybe learning will cease also. It is only if the circle can be broken that we teachers are in a position legitimately to stop providing any more instruction related to the particular point which is thought to have become fossilized. A second point that teachers should bear in mind is the fact that fossilization is not a phenomenon that happens overnight. Although we know of no studies that have specifically investigated this view, it could be said that the curve representing the acquisition of L2 phonology is likely to be similar to the curve of other skills, specially to those which involve some motor control. These curves show that there is very rapid improvement at early stages of learning but that it increases more and more slowly at intermediate stages; finally, it becomes progressively more

stable at advanced stages, so stable, in fact, that it could be assumed that in practical terms a time comes when no further progress is made. This is the stage that can be referred to as fossilization. If the above reasoning is also true of L2 pronunciation acquisition, then it follows that less advanced students would benefit more from a teaching pronunciation programme than would more advanced students, since they are still in a position to make some improvement.

Finally, a methodological point needs mention: that of testing L2 pronunciation accuracy by means of subjective judgments and ratings given by native speakers; many studies are based on these types of measurements (see for example Dimitrijevic and Djordjevic 1971; Mullen 1980; Yorozuya and Oller 1980; Brennan and Brennan 1981; Bezooijen and Hout 1985). In relation to this, in our study we were interested in knowing whether or not there is a basis for the widely accepted belief that linguistically trained people, i.e. linguists and language teachers, are more critical towards the language learner's achievements than are the linguistically naive speakers of that language. If the belief holds any truth, it could be argued that teachers are more severe raters because they are used to hearing a very wide range of 'quality' of pronunciation, which in turn would lead to an implicit comparison of the particular learner with their best students; and even if the learner in question is good they will tend to find some flaws in his or her speech as a result of this comparison. On the other hand, linguistically naive judges, especially if they do not speak other languages themselves, might tend to value the fact that a foreigner speaks their language at all more positively, and therefore tend to be more benevolent in their ratings. In our study we tried to see whether this was true or not.

In more concrete terms, this study was set up to address the following three questions:

1. Is the pronunciation of post-A Level students of Spanish fossilized? By fossilization here we understand no improvement after undergoing a specific Pronunciation Training Programme (PTP).

2. If there is any room for improvement at this level, is there a relationship between amount of quality in pronunciation (that is, how well or badly a particular student pronounces) and amount of improvement (that is, how much measurable progress somebody can make after the relevant instruction)? In our study 'good' and 'bad' pronunciation equals high and low marks on a 7-point scale. 'Amount of improvement' here means amount of 'after' repetitions rated better by the judges.

3. Concerning the use of native judges, are linguistically trained judges harsher in their judgements of foreignness than linguistically naive judges? Harsher judgement here mean giving lower marks on an overall pronunciation 7-point rating scale.

2. Method

2.1 Subjects

The informants of this study were part of a larger set of students who collaborated in the recordings of samples of Spanish as a foreign language. Our subjects were eight native speakers of English (three males and five females) studying Spanish in their first year of an Honours degree in Modern Languages at the University of Newcastle upon Tyne. The average age was 20 and the level of Spanish was post-A Level, which corresponds roughly to upper-intermediate in the level range elementary/lower-intermediate/upper-intermediate/advanced/proficient. All subjects cooperated voluntarily, but their motivation was supposedly quite high because a) it was a self-selected group and b) they had the oral exams for their course coming up shortly after the recordings, and they knew that the study was about pronunciation improvement. The influence that strength and nature of motivation can exert on pronunciation has been emphasized by numerous researchers (Flege 1987a, 1987b, Purcell and Suter 1980, Suter 1976, Gardner and Lambert 1972, Hill 1970).

Other relevant details of the informants' linguistic background were obtained by means of a questionnaire filled in when they first came to the recording studio. Given the multicultural background of some of the students who decide to do a degree in modern languages, it was important to make sure that in our study Spanish was genuinely a foreign language, we wanted a fairly homogeneous group. In fact, some people who had also volunteered for the experiment had to be ruled out on the basis of the information from the questionnaire (either one of the parents was a native speaker of Spanish or they themselves had spent some years in a Spanish speaking country and were virtually bilingual).

The questionnaire was specifically drawn up to meet the terms of the study and consisted of a few very short preliminary questions plus 13 multiple-choice questions about the following aspects related to the experience of learning Spanish (for a full version of the questionnaire we refer the reader to the appendix):

1. Parents' native language.
2. Country in which they spent the first five years of their life.
3. First contact with Spanish: whether it was through formal instruction or through naturalistic exposure, and at what age.
4. Years of formal instruction in the language.
5. Spanish teachers' first language.
6. Time spent in Spanish-speaking country with Spanish-speaking people.
7. Amount of classroom learning addressed to pronunciation of Spanish.
8. Importance attached to pronunciation of Spanish by the learner in regard to effective communication.
9. Concern about their own pronunciation in Spanish.
10. Importance of accurate pronunciation to obtaining a better job.

2.2 Procedure

The study was carried out according to the following procedure:

* selection of speech material and task,
* first recording,
* pronunciation training programme (PTP),
* second recording and
* ratings.

2.2.1 Selection of speech material and task

We used some preliminary speech material which consisted of 130 short model sentences taken from the recordings of Sánchez and Matilla (1986). These sentences had been selected and arranged in sets according to criteria of pronunciation difficulty for English speaking people. The criteria were based on discussions by Sánchez and Matilla (1986) and Stockwell and Bowen (1965), as well as on the researcher's own ideas. Half the sentences were spoken by a man and the other half by a woman, both representative of the standard accent of European Spanish. The semantic content of the sentences was of a neutral kind, of what Gass and Varonis (1984) call 'real world knowledge'. Subjects were to repeat them after only one hearing, so sentences were simple, with no embedding, to minimize the interference of lack of understanding or problems of memory in the repetition.

We decided that repetition served our purposes better than other techniques of eliciting data for pronunciation analysis, such as reading a passage aloud or spontaneous speech. Since we were only concerned with pronunciation and with no other aspect of language - vocabulary, grammar, fluency, etc. - we thought it better if subjects did not have to worry about 'what to say', but only about 'how to say it'. Because we were interested in some quite specific problems of pronunciation, repetition had the further advantage that subjects could not use 'avoidance' as a strategy (Altenberg and Vago 1987; Schachter 1974). We ruled out reading aloud, as this clearly involves skills other than pronunciation, and mistakes may occur as a result of misinterpretation of spelling rather than actual inability to produce a particular sound accurately.

A selection of the material from the first recording (the model sentences repeated by our students) was listened to by three native speakers of Spanish including the researcher herself, and they agreed that there were four groups of consonant sounds primarily responsible for causing 'foreignness', namely the various phonological realizations in Spanish of: 1) the contrast between 't' and 'd' (as in *vengo de parte de Daniel*); 2) 'g' (as in *no me gustan los gatos gordos*); 3) 's' (as in *sus besos me saben a miel*) and 4) the contrast between 'r' and 'rr' (as in *mató al toro de dos tiros, habla un ruso horrible*). These four problematic aspects would constitute the object of our subsequent pronunciation training programme, and were the basis for the selection of material in the second recording, as well as for the final selection of the sentences to be rated by the native judges.

After the first recording a further selection was needed. Clearly it was necessary to remove all the sentences that had turned out to be too long, or too difficult as a result of unknown vocabulary. The rationale behind this further selection was to make sure

that the subject had understood the sentence properly, and that whatever problems he or she might have had in repeating were due to pronunciation difficulties and not to lack of understanding.

2.2.2 Recordings

Subjects were recorded in a recording studio on an individual basis. Each subject was given the same instructions before the researcher left the room. They had to repeat an aural model that was presented to them once; there was a pause of approximately 6-8 seconds between sentences. They could not stop and rerun the tape. It was not possible, with our equipment, to use headphones; we acknowledge the fact that this was a disadvantage, in a task in which repetition depended solely on hearing. Both recordings began with a set of 12 acclimatization-to-the-task sentences, so that subjects could see what the task involved.

The subjects were recorded twice, with a three-month interval and a five-session programme between both recordings:

> First recording: This lasted for about 10 minutes, and consisted of the repetition of the preliminary 130 sentences. Because it was the recording made prior to the PTP, it will also be referred to as the 'before-recording'. Some subjects made comments worth noting here: e.g. about one in three stated that they had found the woman more difficult to understand; some subjects felt that there was not enough time to repeat after some longer sentences; others said that the task was very demanding because it was too long. Although all the problematic sentences were removed from the final set, these aspects might have had a negative effect on the overall performance of some subjects.
>
> Second recording: This lasted for about 5 minutes, and consisted of the repetition of 62 of the above number of sentences, selected on the basis of the four groups of sounds that the study concentrated on. A number of sentences had to be removed because they had proved unsuitable (long, difficult vocabulary...). There was no new material. This recording will also be referred to as the 'after-recording', because it was done after the administration of the PTP.

2.2.3 Pronunciation Training Programme (PTP)

A programme of pronunciation training was designed specifically for the purpose of the study, to focus on the four pronunciation problems of our students. It consisted of five one-hour sessions held over ten days two weeks before the second recording; every session was divided into two parts:

1. Use of drama techniques.
2. Practice exercises in the language laboratory.

2.2.3.1 Use of drama techniques

Drama techniques have been widely applied in second language teaching, especially in the teaching of spoken communication skills and pronunciation. The particular techniques employed in our sessions are closely related to those practised by actors

in the theatre. The underlying principle is that to make full use of the voice, one needs to control all aspects involved in oral production: shaping of the mouth, posture, the mechanics of breathing, facial muscle control, etc. Our point here is that these aspects are even more significant when we pronounce in a foreign language. It is difficult enough to have to impose new articulatory habits on the ones that the learner has been operating with for so many years. The picture is worsened if we consider that speaking a foreign language has something about it of a 'public performance' - very much like actors on stage. The learner is usually too tense; the whole of the articulatory apparatus becomes stiff and rigid and therefore the free movement of tongue, lips and jaw is highly constrained.

Exercises on relaxation and posture, breathing, tone and articulation, have proved most useful in teaching L2 pronunciation. For a detailed account of how these techniques work in practice, we refer the reader to Wessels (1987). Many of the activities and exercises of our pronunciation programme were derived from this book as well as from various drama sessions conducted in Edinburgh in 1988 by the author and attended by this researcher.

2.2.3.2 Practice exercises in the language laboratory

The preparation of the exercises in the pronunciation training programme (PTP) were based on the following ideas:

1. Working on precise segmental problems which carry heavier weight in Spanish 'foreignness'.

2. Explaining and contrasting the differences between corresponding phonemes of the native and the target language.

3. Repeating and rehearsing (Asher and Garcia 1969) helps to produce a better performance. The idea is that practice makes, if not perfect, at least better (Tahta et al. 1981a; Tahta et al. 1981b).

4. One of the key principles of the PTP is the administration of immediate feedback of the performance by way of comparing one's pronunciation with that of the native speaker.

The PTP was aimed at correcting errors in the various pronunciations of Spanish 't' versus 'd', 'g', 's' and 'r' versus 'rr'. We considered that some work on vocalic sounds - especially on diphthongs - was also advisable. It goes without saying that at no time was the PTP thought to cover all pronunciation problems that our students presented. All sessions were conducted in Spanish by the researcher.

2.2.4 Ratings

2.2.4.1 Raters

Eight native speakers of Standard European Spanish took part in rating the performance of our subjects. Four raters had a linguistic background or were themselves teachers of Spanish as a foreign language. The other four were linguistically naive judges and had no teaching experience of any kind.

2.2.4.2 Rating material

After the two recordings, 10 of the model sentences were selected for the analysis. The list of sentences can be found in the Appendix. Each model sentence had two realizations by each subject, one from the first recording and one from the second recording. Once the whole of the relevant material had been decided on we had to 'prepare' it appropriately for the raters. This is how it was arranged:

1. The 10 native models of the sentences were recorded on to a separate tape that we shall call the MASTER set.

2. For every subject the two relevant realizations ('before' the PTP and 'after' the PTP, pairs always in this order) of the 10 model sentences were identified and re-recorded in the same order as in the master set.

3. Next, for every subject's material two processes of randomization were carried out:

   3.1. Inter-pair randomization. The order of the 10 pairs, taken as 10 units, was randomized.

   3.2. Intra-pair randomization. The 'before' and 'after' PTP order of every pair was randomized, so that raters would not know *a priori* if they would be listening first to the 'before' or the 'after' realization.

4. The whole of the rating material was then recorded again. In the new recording, every pair of realizations was preceded by the corresponding native model. The recordings were arranged in two separate sets: 1) a FIRST SET, containing the speech material from subjects 1, 2 ,3 and 4; and 2) a SECOND SET, containing the speech material from subjects 5, 6, 7 and 8. By keeping these two sets separate we were trying to neutralize the effect of fatigue on the work of the raters; so, 4 raters would listen to the FIRST SET (subjects 1, 2, 3 and 4) first, and the other 4 raters would listen to the SECOND SET (subjects 5, 6, 7 and 8) first.

5. Two more tapes were also edited. The first one we called the WARM-UP set: it contained five pairs (the first pair of subjects 1, 2, 3, 4 and 5) and its function was to acclimatize the judges to the rating task, making sure they had understood the instructions. We called the second tape the CONSISTENCY set: it also contained five pairs (the first pair of subjects 4, 5, 6, 7 and 8) and its purpose was to check whether raters were consistent in their judgements.

6. In short, there were five different sets of material for the judges to listen to:

   * The MASTER set;
   * the WARM-UP set;
   * the FIRST set (or SECOND set);
   * the SECOND set (or FIRST set);
   * the CONSISTENCY set.

2.2.4.3 Rating tasks

All raters did their rating alone with the researcher. They were given written instructions (Appendix C) that they had to read carefully. As part of the instructions raters listened to the MASTER set, to acquaint themselves with the 10 model sentences and the voices of the native speakers. They also listened to the WARM-UP set, to become familiar with the rating sheet and the rating task itself. Attention was drawn to two important aspects of the material:

- The volume of voice of some speakers was on occasions a little 'low' (mainly for technical reasons or shyness of the speaker). Judges were urged to bear this in mind and not to let it influence the rating.
- Hesitations or repetitions of words should not be regarded as flaws in the quality of pronunciation itself.

Judges had to do two different kinds of rating task. The overall rating time was about 35 to 40 minutes. They could not stop and rerun the tape. There was a 2-3 minute break between each set.

A) First rating task

See rating sheet in the appendix. For every subject, pairs had been numbered 1 to 10. This first column matched three other columns containing the options: 'first repetition', 'second repetition' or 'Don't know'. Judges had to listen to the model and then the two repetitions and decide which of the two was better and tick the appropriate column. They had been urged always to make a judgement, and only to tick the 'Don't know' column if they really could not make up their minds as to which had better pronunciation. Listening to and rating the CONSISTENCY set was also part of this first task.

B) Second rating task

After completing the first task, judges additionally provided a separate rating of the overall pronunciation of every subject on a seven-point scale ranging from 1 (completely unintelligible pronunciation) to 7 (native accent).

2.3 Analysis of results

This study addressed three questions and we shall deal with them in turn.

2.3.1 Is the pronunciation of post-A level students of Spanish fossilized?

Results are summarized in table 1.

Table 1 *First task rating results for every subject*

Subjects	'Before' better	'After' better	'Don't know'	$\chi 2$	Significance ($p < .05$)
1	27	46	7	2.47	not sig.
2	37	35	8	0.028	not sig.
3	51	18	11	7.89	sig. worse
4	22	46	12	4.24	sig. better
5	19	55	6	8.76	sig. better
6	50	22	8	5.44	sig. worse
7	21	47	12	11.67	sig. better
8	35	37	8	0.028	not sig.
TOTAL	262	306	72	1.704	not sig.

Column 1: subjects; column 2: total number of 'before' repetitions rated as better; column 3: total number of 'after' repetitions rated as better; column 4: number 'Don't knows'; column 5: χ^2 value of the differences between 'before' and 'after'; column 6: significance.

If the null hypothesis was true, then there would be no effect of the PTP on the 'after' repetitions of the students (i.e. their pronunciation was fossilized).

We obtained from the judges 640 responses altogether - distributed in 'befores', 'afters' and 'don't knows'. The number of 'don't knows' was removed from the calculations on the grounds that when a judge was not sure which of the two realizations was better he or she was actually withholding judgement and these answers could not be counted either way.

From this analysis we can conclude that our null hypothesis holds, that is, no significant improvement was observed in the pronunciation of our students after the PTP. Let us see what happened when we considered not the group as a whole but the subjects separately:

1. Three subjects (4, 5 and 7) did significantly better in the 'after' recording.

2. Three subjects (1, 2 and 8) did not show any significant difference.

3. Two subjects (3 and 6) did significantly worse in the 'after' recording. This third point suggests that, for some students at least, the PTP was not only of no value at all but in fact had had some negative effect on their performance. This is an unexpected result and we shall come back to it in our discussion.

2.3.2 Did weaker students improve more than better students?

In the answer to our previous question we saw that some students had improved and some had not. What can we then say about the relationship between the amount of improvement and the quality of the subjects' pronunciation? To find out we calculated a Spearman Rank Correlation Coefficient (r_S). It has already been explained that every subject was rated eight times on his or her overall pronunciation on a seven-point scale ranging from 1 (completely unintelligible accent) to 7 (native accent). We could now rank our students according to the two variables: the average score of how good or bad our raters thought they were and the number of better 'after' repetitions. Results are shown in table 2.

Table 2. *Ranking of subjects according to overall pronunciation.*

Subjects	Overall pronunc.	improvement after PTP	Pronunc. rank	Improvement rank
4	3.34	46	8	5.5
7	3.56	47	7	7
7	3.81	55	6	8
1	4.65	46	5	5.5
8	4.75	37	4	4
6	4.96	22	3	2
2	6.28	35	2	3
3	6.62	18	1	1

The table shows the relationship between quality of overall pronunciation and amount of improvement after the PTP.

From the table it seems that the lower the rating in overall pronunciation the more likely the subject is to gain improvement from the PTP. We tested our hypothesis of no correlation by carrying out a Spearman Rank Correlation Coefficient. Our value of $r_S = 0.851$ indicates a significantly ($p < .05$) strong negative correlation between our two variables, that is, the worse the pronunciation the greater the improvement from the PTP.

2.3.3 Are linguistically trained native speakers stricter in rating foreign students' pronunciation than linguistically naive native speakers?

To find this out we had to compare the ratings that the two different groups of raters had given to every student (table 3). The comparison was made by carrying out a Wilcoxon Matched-Pairs Signed-Ranks Test. This non-parametric measure makes no assumptions about the distributions of the ratings and is perhaps a safer statistic to use than a paired t-test. For every subject we had two mean values for overall pronunciation - one given by linguistically trained raters and one by linguistically non-trained raters.

Table 3. *Comparison of means obtained in ratings made by linguistically trained (LT) and linguistically non-trained (LNT) native raters.*

Subjects	LT Rating Means	LNT Rating Means
1	4.2	4.75
2	6.25	6.31
3	6.44	6.81
4	3.12	3.56
5	3.87	3.75
6	5.05	4.87
7	3.44	3.69
8	4.75	4.75

We obtained a value for T of 5 (critical value of T for significance at 5% level is 2) and thus we have to conclude that there was no difference between the way that the linguistically trained and linguistically naive native speakers rated pronunciation.

3. Discussion and further questions

From the analysis it appears that as a group our students made no significant improvement after the PTP. At the same time is clear that some students significantly benefited from it. There seems to be some confirmation of a tendency to show more phonological fossilization at more advanced levels of the language. With regard to judgements of foreignness, results seem to contradict the idea that some people (particularly students) have that teachers judge more strictly than other interlocutors.

3.1 How far can we generalize our results?

We are fully aware of the two main limitations of this study: the small number of subjects and the small number of sessions of the PTP. Our results should be taken rather as an invitation to further research. We do not know, for example, if a PTP twice as long will have better effects on the degree of improvement. Besides, is this improvement a lasting effect or does it fade away shortly after the second recording? Furthermore, we cannot at this stage be certain that the students who showed a significant improvement in controlled speech would really do better in spontaneous conversation.

3.2 How sure can we be that improvement in pronunciation was due to the PTP and not to an uncontrolled variable?

This question takes the explanation of improvement a step further. One such variable could be consciousness raising: the pronunciation of some of our students could have improved simply because attention had been drawn to the existence of problematic areas, irrespective of the PTP. We think, however, that consciousness raising is not a distinct variable but rather part of the explanation why a pronunciation programme may bring about some progress, because of the very nature of any such programme.

A second factor that may account for the improvement in pronunciation is a desire on the part of the subjects to please the experimenter. We think that the effect of this factor in our study may only be negligible, because even if it had had some influence on the performance it would have probably affected both recordings equally, and therefore it would not have biased the relative results of our data.

3.3 Why were there two students that did significantly worse after the PTP?

This is a surprising result that needs looking into carefully. It is difficult to believe, although it might just be possible, that a PTP causes harm to somebody's pronunciation, unless, that is, the guidelines contained in the exercises for practice were utterly wrong. We are very confident that this was not the case in our PTP, because it had been meticulously planned according to a well documented description of the Spanish phonetic system (Navarro Tomás 1982; Stockwell and Bowen 1965).

Instead, we suggest two possible explanations for this unexpected negative effect. The first one is of a technical nature, and it refers to a difference in the volume at which the two recordings were made. There is evidence that this could well have been the case with the subject that showed the highest negative effect of the PTP (subject 3). For reasons that we could not control, the volume of his after-PTP recording was noticeably lower than that of the first recording. Although we had warned our judges about this technical problem it is possible that they consistently gave the 'better' score to the realization which they heard with less difficulty. This explanation is supported by the fact that both students in question had very good pronunciation according to the judges and, therefore, it was unlikely that there would be much difference between the two realizations of the sentences other than the actual volume of the recordings. This result underlines how important it is for the subsequent analysis of data to obtain as high-quality recordings as possible.

The second possible explanation that we would like to consider relates to a much more significant aspect. It suggests that when an individual has acquired a high standard in non-monitored speech, as was the case with these two subjects, concentration on very controlled items of speech may lead them to pronounce somehow 'less naturally' and therefore 'less authentically'. It this is true, we could further suggest that the more such student concentrates on a particular item, the less natural he or she may sound. It seems logical to conclude that the second recording sounded 'less natural', subjects 'overdid it', because the PTP had made them concentrate on the speech material even more than on the first occasion.

4. Conclusions: implications for teaching

In our previous sections we have discussed some problems related to the learning of pronunciation. If research is to have any value in practical terms, findings need to shed some light on the everyday practice of second language teaching. It is very important that teachers should be convinced that achieving a good pronunciation is something that needs attention on their part, which in practice means attention to the mistakes of the students and their correction.

Acquiring a good pronunciation is not something that occurs automatically, but it is a process that can be helped. MacCarthy (1978) points out the enormous help that

teachers can offer students before they go to a foreign country for their university year abroad. Being permanently immersed in the target language speaking situation does not automatically lead to a near-native command of the language. It is the role of the teacher to 'open' the students' ears and make them conscious of which particular aspects make 'native' pronunciation sound more native. This means working on auditory training before students are sent off to the foreign country. It is these less advanced students that are most likely to benefit from some kind of pronunciation training, as suggested by the results of our study.

Purcell and Suter (1980) argue that teachers and classrooms seem to have remarkably little to do with how well our students pronounce, since variables other than formal instruction seem to have much more impact on predicting pronunciation accuracy. However, we think that a good teacher can still stimulate the increase of concern and motivation that Purcell and Suter regard as so important for acquiring a good accent.

In the case of weak students, working on pronunciation can help them gain some confidence; this in turn may make them more active in linguistic interactions. Teachers may start by raising an overall awareness towards the spoken language. One of the aims, then, is to train students to listen actively, to 'observe' the target language. Good pronunciation is not only about uttering beautifully articulated sentences, it presupposes a high degree of ability to detect small acoustic differences.

In learning good pronunciation there is probably an element of 'naturalness' combined with an element of 'awareness'. It is our view that it is the latter that teachers must concentrate on to try to help students to achieve a better pronunciation. Neufeld (1977) has shown that adults can learn very good pronunciation, and this finding should encourage both teachers and researchers to find more ways of dealing with the remaining problems concerning the acquisition of the phonology of a foreign language.

References

Altenberg E. and R. Vago. 1987. 'Theoretical implications of an error analysis of second language phonology production' in Ioup and Weinberger (eds.) 1987:148-164.

Asher J and R. García. 1969. 'The optimal age to learn a foreign language'. Modern Language Journal 53: 334-341.

Bezooijen R. van and R. van Hout 1985. 'Accentedness ratings and phonological variables as measures of variation in pronunciation'. Language and Speech 28/2: 1985.

Brennan E. and J. Brennan 1981. 'Accent scaling and language attitudes: reactions to Mexican American English speech'. Language and Speech 24/3: 207-221.

Brière E. 1966. 'An investigation of phonological interference'. Language 42/4: 768-796.

Dimitrijevic N. and D. Djordjevic 1971. 'The reliability of subjective assessment of pupils' pronunciation of English as a foreign language'. IRAL 9/3: 245-265.

Flege J. 1980. 'Phonetic approximation in second language acquisition'. Language Learning 30/1: 117-134.

Flege J. 1981. 'The phonological basis of foreign accent: a hypothesis'. TESOL Quarterly 15/4: 443-455.

Flege J. 1987a. 'Limits on phonetic accuracy in foreign language speech production' in Ioup and Weinberger (eds.) 1987: 176-203.

Flege J. 1987b. 'A critical period for learning to pronounce foreign languages'. Applied Linguistics 8: 162-177.

Gardner R. and W. Lambert 1972. Attitudes and Motivation in Second Language Learning. Rowley, Mass: Newbury House.

Gass S. and E. Varonis 1984. 'The effect of familiarity on the comprehensibility of non-native speech'. Language Learning 34/1: 65-89.

Guiora A., B. Beit-Hallahmi, R. Brannon, C. Dull and T. Scovel 1972. 'The effects of experimentally induced changes in ego states on pronunciation ability in a second language: an exploratory study'. Comprehensive Psychiatry 137: 421-424.

Harder P. 1980. 'Discourse as self-expression. On the reduced personality of the second language learner'. Applied Linguistics 1/3: 262-270.

Hecht B. and R. Mulford 1987. 'The acquisition of a second language phonology: interaction of transfer and developmental factors' in Ioup and Weinberger (eds.) 1987: 213-227.

Hill J. 1970. 'Foreign accent, language acquisition and cerebral dominance revisited'. Language Learning 20: 237-248.

Ioup G. and S. Weinberger (eds.) 1987. Interlanguage Phonology. The Acquisition of a Second Language Sound System. Cambridge, Mass: Newbury House.

Krashen S. 1973. 'Lateralization, language learning and the critical period: some new evidence'. Language Learning 23/1: 63-74.

Krashen S. 1985. The Input Hypothesis. Issues and Implications. London: Longman.

Loewenthal K. and D. Bull 1984. 'Imitation of foreign sounds. What is the effect of age?'. Language and Speech 27/2: 95-98.

MacCarthy P. 1978. The Teaching of Pronunciation. Cambridge: Cambridge University Press.

Mullen K. 1980. 'Rater reliability and oral proficiency evaluations' in J. Oller and K Perkins (eds.) 1980. Research in Language Testing. Rowley, Mass.: Newbury House. 91-101.

Navarro Tomás T. 1982. Manual de Pronunciación Española. Madrid: CSIC.

BEST COPY AVAILABLE

Neufeld G. 1977. 'Language learning ability in adults: a study on the acquisition of prosodic and articulatory features'. Working Papers on Bilingualism, University of Ottawa 12: 47-60.

Purcell E. and R. Suter. 1980. 'Predictors of pronunciation accuracy: a reexamination'. Language Learning 30/2: 271-287.

Sánchez A. and J. Matilla. 1986. Manual de Corrección Fonética del Castellano. Madrid: SGEL.

Schachter J. 1974. 'An error in Error Analysis'. Language Learning 24/2: 205-214.

Schumann J. 1976. 'Affective factors and the problem of age in second language acquisition'. Language Learning 25/2: 209-235.

Scovel T. 1969. 'Foreign accents, language acquisition and cerebral dominance'. Language Learning 19/3 & 4: 245-254.

Selinker L. 1972. 'Interlanguage'. IRAL 10/2: 209-231.

Snow C. and M. Hoefnagel-Hohle. 1982. 'The critical period for language acquisition: evidence from second language learning' in Krashen S., R. Scarcella and M. Long (eds.) 1982. Child-Adult Differences in Second Language Acquisition. Rowley, Mass.: Newbury House.

Stockwell R. and J. Bowen. 1965. The Sounds of English and Spanish. Chicago: Chicago University Press.

Suter R. 1976. 'Predictors of pronunciation accuracy in second language learning'. Language Learning 26/2: 233-253.

Tahta S., M. Wood and K. Loewenthal. 1981a. 'Age changes in the ability to replicate foreign pronunciation and intonation'. Language and Speech 24/4: 363-372.

Tahta S., M. Wood and K. Loewenthal. 1981b. 'Foreign accents: factors relating to transfer of accent from the first language to a second language'. Language and Speech 24/3: 265-272.

Tarone E. 1987. 'The phonology of interlanguage' in Ioup and Weinberger (eds.) 1987: 70-85.

Wessels C. 1987. Drama. Resource Books for Teachers. Oxford: Oxford University Press.

Wode H. 1976. 'Developmental sequences in naturalistic second language acquisition'. Working Papers in Bilingualism 11: 1-31.

Yorozuya R. and J. Oller. 1980. 'Oral proficiency scales: construct validity and the halo effect'. Language Learning 30/1: 135-153.

Appendix
Questionnaire

INSTRUCTIONS

Here is a questionnaire about your experience learning Spanish. You will be asked questions about specific facts of your life as well as your personal attitude to the problem of pronunciation in a second language. In every question you will see a number of statements. Please, read them carefully and then tick the box that most accurately reflects your own experience or feelings in every case.

NAME _____ AGE _____ SEX _____
YOUR NATIVE LANGUAGE IS _____
YOUR FATHER'S NATIVE LANGUAGE IS _____
YOUR MOTHER'S NATIVE LANGUAGE IS _____
COUNTRY IN WHICH YOU SPENT THE FIRST FIVE YEARS OF YOUR LIFE _____
YOUR COURSE AND YEAR AT UNIVERSITY _____

1 How did you first come into contact with Spanish?

[ ] Through formal instruction in an English-speaking environment
[ ] Through formal instruction in a Spanish-speaking environment
[ ] Through 'naturalistic' exposure in a Spanish-speaking country
[ ] Other way (please specify)_____

2 When did you first start learning Spanish?

[ ] Before the age of 7
[ ] Between the ages of 7-12
[ ] Between the ages of 12-16
[ ] After the age of 16

3 How many years of formal instruction in Spanish have you had?

[ ] Less than 3 years
[ ] 3 to 5 years
[ ] 6 to 8 years
[ ] More than 8 years

4 Your Spanish teachers

[ ] All of them have been native Spanish speakers
[ ] All of them have been native English Speakers
[ ] Most of them have been native Spanish speakers
[ ] Most of them have been native English Speakers

5 How long altogether have you spent in a Spanish speaking country with Spanish speaking people?

[ ] Less than 1 month
[ ] Between 1 month and 6 months
[ ] Between 6-12 months
[ ] Between 12-24 months
[ ] More than two years

6 How much formal classroom training directed specifically to pronunciation of Spanish have you had?

[ ] I have had no specific training at all
[ ] I was only given some guidance at the beginning of my learning
[ ] I have followed a course or done some specific practice periodically

7 To communicate effectively in Spanish, you think you need

[ ] To pronounce exactly as as native speaker would
[ ] To pronounce more o less correctly
[ ] To speak intelligibly

8 In relation to grammar and vocabulary of a second language you think that

[ ] Pronunciation is less important
[ ] Pronunciation is as important as grammar and vocabulary are
[ ] Pronunciation is more important

9 During the time you have been learning Spanish, do you think that your concern about pronunciation has been

[ ] Less than your course mates
[ ] About the same as your course mates
[ ] Greater than your course mates

10 After a number of years [illegible]arning a second language, do you think that pronunciation

[ ] Can still improve, as can grammar and vocabulary
[ ] Can still improve, but to a lesser extent than grammar and vocabulary
[ ] Cannot improve

11 To improve your pronunciation in Spanish after a number of years, do you think you should mainly

[ ] Listen to native speakers
[ ] Talk to native speakers
[ ] Listen to yourself on tapes
[ ] Receive formal instruction on pronunciation

12 When you converse in English with a non-native speaker of English, how do you feel about bad pronunciation?

[ ] It will not bother you as long as you can understand what he is saying
[ ] It will certainly put you off the conversation
[ ] You would think twice before engaging in a second conversation with that person

13 Do you think that you stand a better chance to obtain the kind of job you want if you pronounce Spanish well?

[ ] Yes, I think so
[ ] I am not sure
[ ] The accuracy of my Spanish pronunciation is irrelevant to the sort of job I want

THANK YOU VERY MUCH

Model sentences

1. Vengo de parte de Daniel
2. Mañana me lo darás todo
3. No me gustan los gatos gordos
4. Gota a gota se llena la bota
5. Sus besos me saben a miel
6. Esta especialidad no es para mí
7. María no quería mirarte a la cara
8. Mató al toro de dos tiros
9. Me gustan esas ropas rojas
10. Habla un ruso horrible

101

RATING SHEET

DATOS PERSONALES

NOMBRE _____

Profesor de español Si NO Sexo H M

Region de procedencia _____

Edad Años

Prueba nº 1			
		Mejor	
		Pronunciac.	
Par.	1a	2a	No se
1			
2			
3			
4			
5			

Codigo ___			
Sujeto nº ___			
		Mejor	
		Pronunciac.	
Par.	1a	2a	No se
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Codigo ___			
Sujeto nº ___			
		Mejor	
		Pronunciac.	
Par.	1a	2a	No se
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Codigo ___			
Sujeto nº ___			
		Mejor	
		Pronunciac.	
Par.	1a	2a	No se
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Codigo ___			
Sujeto nº ___			
		Mejor	
		Pronunciac.	
Par.	1a	2a	No se
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Codigo ___			
Sujeto nº ___			
		Mejor	
		Pronunciac.	
Par.	1a	2a	No se
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Codigo ___			
Sujeto nº ___			
		Mejor	
		Pronunciac.	
Par.	1a	2a	No se
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Codigo ___			
Sujeto nº ___			
		Mejor	
		Pronunciac.	
Par.	1a	2a	No se
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Prueba nº 2			
		Mejor	
		Pronunciac.	
Par.	1a	2a	No se
1			
2			
3			
4			
5			

BEST COPY AVAILABLE

INTERPRETING METONYMY

Anne Pankhurst (DAL)

Abstract

This paper discusses some of the problems of interpreting metonymy, a commonly used but little understood trope. Initial clarification of the differences between metonymy, synecdoche and metaphor is proposed, using common examples. It is suggested that an account of conventionalised and creative metonyms, and the underlying principles common to them, must include both descriptive and analytical considerations. A relevance-theoretic framework is investigated with a view to finding further explanatory insights. Exemplification of the role and functions of metonymy in narrative fiction is presented in texts by Charles Dickens and Virginia Woolf. The outcome of the study shows that rhetorical, linguistic and relevance-theoretic frameworks are all needed if metonymy, and the process of interpreting its meaning, is to be fully accounted for.

1. Introduction

1.1 Definitions of metonymy.

Metonymy is frequently found in normal, everyday language, particularly in conventionalised examples such as:

Have you done your Chomsky? (the author for the work).
The strings played superbly (instruments for agents).
We'll have the Bordeaux (area of production for product).
The White House has refused (place for the authority).
That green anorak is coming (garment for wearer).
We could do with some fresh blood (part of body for the person).
Small is beautiful (general for specific).

Various attempts (Schofer and Rice 1977, Lakoff and Johnson 1980, Nash 1989) to describe metonymy provide lists like this, creating a taxonomy of examples. This categorising fails to explain why the trope is used, its functions in discourse and how it achieves effects. Metonymy is generally not well known, or easily identified, in spite of its being in common use. It is frequently confused with two other common figures of speech, metaphor and synecdoche, as in, for example, the non-literal use of the word *ship*.

Firstly, as metonym, *ship* can stand for a place in which people live and work over a period of time, signifying a large group of people found together in one ship as in *The entire ship rejoiced*. The referent in spatio-temporal location may thus be extended

from the scene to the characters of a narrative. Metonymy is generally agreed to mean the substitution of one term for another, on the basis of some material, causal or conceptual relationship between terms within a single domain (Preminger 1975). It enables the extension of a physical reference into more abstract connotations. Hence, *ship* acquires different connotations associated with the emotions or actions of the people aboard. The closely related figure of speech synecdoche is generally seen as a special kind of metonymy (part for the whole) and for the purposes of this paper will be conflated with it. In synecdochic description, a part of a ship can stand for the whole but within the limits of physical features as in *There is a sail entering harbour*.

Secondly, as metaphor, *ship* can be substituted for a concept such as the State, or an organisation such as a business, an educational institution, e.g. *The ship of state needs the right person at the helm*, and *Our chairman runs a tight ship*. The actions associated with a ship may be part of a generally accepted conceptual metaphor 'life is a journey' (cf. Lakoff and Johnson 1980) as in *sailing through stormy seas*. Metaphor is a substitution of one term for another on the basis of an analogy between the two terms, which come from different domains. In this example, the domain of 'life' is transferred to the domain of 'transport by ship' which it is perceived to resemble.

1.2 Some principles of metonymy.

The example of *ship* demonstrates that in the case of both metaphor and metonymy there is a gap between the linguistic form and the semantic meaning of the word or phrase. Attempts to separate metonymy from metaphor raise interesting questions about the boundaries between the two tropes. The discussion in this section of the paper will, however, focus on the generally agreed principles and uses of metonymy.

Metonymy is primarily referential in character, relying on causality as a linking principle between the term and its referent, which is in the material world (Schofer and Rice 1977). Another shared principle is contiguity, essentially spatio-temporal contact, primarily in conventionalised figures such as place for person (e.g. *The White House has refused*) but extended to any shared linguistic or extra-linguistic code (Jakobson 1956), or semantic associations (Eco 1979). This kind of contiguity is context-linked, as can be seen in the example *Have you done your Chomsky?* which is not perceived as meaningful outside its specific environment, an Applied Linguistics course.

In some metonyms a kind of ellipsis is used to achieve semantic effects. For example, *We'll have the Bordeaux* is easily understood by anyone who knows that Bordeaux is a kind of wine produced in a specific area, and that the speakers are probably in a restaurant. In a shared context, the full literal version such as *We'll have the wine from Bordeaux* is felt to be unnecessary. Creative examples are seen in political or advertising slogans in which inference plays a considerable part. *Small is beautiful* carries more attention-taking effect than a literal version of the same proposition such as *It is desirable to plan for small-scale development*. The reader infers meaning through his knowledge of the socio-economic code.

The principle of understanding through a process of inference brings us to a relevance-theoretic explanation of figurative language (Sperber and Wilson 1986b). Metonymy is no more than a means of communicating the speaker's intention, and is

understood through the same principle (inference) as other figures of speech such as metaphor, hyperbole and irony. The brief examples given, however, are largely decontextualised, and the explanation does not account for special effects derived from using metonymy. The commonly used metonymy of clothes for wearer, as in *That green anorak is coming* (above) has a range of effects recoverable only from knowledge of the immediate context of the utterance - feelings about the colour, the garment, its wearer, shared opinion etc.

In this paper I shall discuss whether adding a relevance-theoretic account of metonymy to traditional rhetorical and linguistic accounts provides a more adequate explanatory framework. In two examples taken from narrative fiction I shall use a textual analysis as a basis from which to assess whether applying relevance theory makes a contribution to understanding the process of interpretation. In doing so, I shall consider the effects achieved by metonymic discourse (longer text whose organising principle is metonymic) and consider whether these are different from those found in short examples.

2. Relevance Theory

2.1 The principle of relevance.

Relevance theory is a theory of verbal communication developed with the purpose of accounting for the meaning of utterances. It is claimed (Sperber and Wilson 1986a) that the principle of relevance is an explanatory generalisation which applies without exception to all ostensive-inferential acts of communication. Participants in overt acts of communication do not need to be consciously aware of 'rules' of relevance in order for relevant communication to take place. The assumption is that since human cognition is relevance-oriented, there is an expectation that every utterance is relevant. The speaker is presumed to intend an utterance to be relevant; the hearer is presumed to know this before interpretation begins.

The combination of contextual effect and processing effort is fundamental to relevance theory. Contextual (or cognitive) effects are changes in the shared cognitive environment, brought about by the speaker's intention and the inferences drawn by the hearer. They may be derived from different types of knowledge, lexical, logical or encyclopaedic. Achieving contextual effect, or sharing cognitive knowledge through the act of communication, is linked to the amount of effort required of the hearer as he interprets meaning.

Since the degree of relevance depends on the two factors of effort and effect, it increases if less effort brings greater effect, and decreases if more effort brings less effect. Processing effort accesses a range of strong and weak implicatures, which are crucial if optimal relevance is to be achieved. If an utterance represents literal truth or a conventional use of figurative language, we may speak of strong implicature. When considering creative figurative language, however, a wide range of weak implicatures is accessed. More processing effort is required in the latter case, bringing commensurably greater cognitive effects.

2.2 Optimal relevance.

Wilson and Sperber (1993) claim that all ostensive-inferential communication is governed by a principle of optimal relevance shared by speaker and hearer. It goes like this: an optimally relevant utterance must be worth the hearer's attention, and put the hearer to no gratuitous processing effort in achieving its effects. Relevance is not solely the responsibility of the speaker. The hearer, who expects relevance, plays an active part in resolving difficulties caused by ambiguity, linguistic encoding or context. In other words, linguistic and encyclopaedic knowledge interact during interpretation, whether the expression is literal or figurative.

A more controversial aspect of the principle of optimal relevance is that the first satisfactory interpretation is the only satisfactory interpretation, and is the one the reader chooses. Gratuitous additional processing is said to reduce the relevance of an utterance. In the interests of achieving their communicative goal, writers frame their text in such a way that the first plausible interpretation reached by the reader is the one they wanted to convey. According to the principle of least effort, it is assumed that the reader stops when he reaches what he believes to be the intended meaning. At this point, a satisfactory range of intended effects will have been reached although not necessarily all those available. In the case of figures of speech, a more creative reader will have to put more effort into disambiguation and contextualisation of referents which are not readily accessible. His reward will be that he will gain more effects, and more relevance.

In the process of reference assignation, disambiguation of linguistic encoding and the need for contextual enrichment may present difficulty for the interpretation of an utterance. The suitability of referents is evaluated in a certain order by the hearer or reader, starting with the familiar and proceeding to the less familiar. Although this is a generalisation, there is particular interest here for the case of metonymy. When metonymy is used creatively, a great deal of its effect derives from unexpected deviance from literal truth, or the 'gap' between linguistic form and semantic implication, which arouses the reader's curiosity and invites him to search for a less obvious referent. This is particularly true of narrative fiction but not exclusive to it. It has been noted in examples such as emblematic names e.g. *The Iron Lady*. The willingness of the reader to be imaginative in responding to the writer is essential if the text is to make its full effect.

2.3 Interpretive resemblance.

We recall that all tropes depend on a relationship between terms. Metonymy depends on material, causal or conceptual relationships. It is one of the possible representations of the connections between thought and linguistic expression, albeit one which is not always transparent. Sperber and Wilson (1986a) suggest that a relationship between terms may be either descriptive - representing in virtue of truth conditions; or interpretive - representing in virtue of resemblance. They call the latter *interpretive resemblance*, using this term to explain all figurative language. Shared implications explain the relationship in any figure of speech such as metonymy, metaphor, hyperbole, etc. where there is a clear difference between the form of the utterance and its pragmatic implications. The reader assumes that there are some

identifiable characteristics in common, and the same interpretive abilities and procedures are invoked as in the interpretation of literal expressions.

The relationship between the form of a conventionalised metonymic utterance and its referent requires relatively little interpretive effort because of the highly and often specifically referential character of metonymy. There is a close resemblance between the expression and the implicature, as may be seen in the frequent metonymic use of clothing to stand for the person. Although the clothes do not literally resemble the person, they are closely linked by logical and conceptual ties. In the case of an example such as *That green anorak is coming*, constraints on the mobility of an anorak enable prompt disambiguation and access, through inference, to the implicature, that the speaker has some kind of attitude about the wearer of the anorak, which she wishes to convey. An interesting question is whether this is true for longer text in which metonymy is a structuring principle whose effect lies in the accumulation of, say, descriptive detail, or is used creatively to achieve distinct poetic effects.

3. Exemplification of metonymy in literary texts.

I propose to discuss the interpretation of metonymy in the light of exemplification from narrative fiction, using rhetorical and linguistic analysis, and to follow this with considerations based on relevance-theoretic principles. The examples chosen represent different styles of fictional writing and different ways of using metonymy as a structuring principle.

3.1 Dickens: *Bleak House*

The opening pages of Dickens' novel *Bleak House* are frequently cited as an example of metonymic discourse in fiction. By metonymic discourse is meant discourse which uses metonymy - or substitution of one term by another in close material, causal or conceptual relationship - as a fundamental structure for conveying meaning. The novel abounds in examples, including frequent use of synecdochic detail, so that metonymy is a structuring principle throughout the narrative. It is part of the linguistic and conceptual framework which enables interpretation.

In this text, the implied narrator is George Rouncewell, a family servant, who is patrolling Sir Leicester Dedlock's London mansion in the early hours of a snowy morning after Lady Dedlock's disgrace and flight. His gloomy premonitions of disaster are supported by his observations.

> There is no improvement in the weather. From the portico, from the eaves, from the parapet, from every ledge and post and pillar, drips the thawed snow. It has crept, as if for shelter, into the lintels of the great door - under it, into the corners of the windows, into every chink and crevice of retreat, and there wastes and dies. It is falling still; upon the roof, upon the skylight, even through the skylight, and drip, drip, drip, with the regularity of the Ghost's Walk, on the stone floor below.
>
> (Dickens, *Bleak House* Ch.58:855)

A wealth of realistic detail has, by this stage of the novel, informed the reader of scenes, plots and characters, but runs the risk of confusing him by its sheer mass. The problem for the reader is to decide which elements of the descriptions and narration are important enough to be selected from among the rest. It is known that one of the ways in which Dickens marks themes is by using figurative language. He foregrounds important ideas by a combination of tropes and syntactic variations central to the development of given and new information, and leading to the creation of rich contextual effects.

In the immediate context of the most recent parts of the narrative, meaning is achieved by grounding potentially figurative effects in the reality of cold, invasive weather. This is generally agreed to be a feature of metonymic description and is based on the gap perceived to exist between the syntactic or lexical form and the semantic implications, which has to be filled by inference. The first clue to this interpretation is evident from the beginning of the novel, in the description of the London fog. With this established as a frame, the state of the weather functions as a clue to the reader to draw inferences. The reader's encyclopaedic knowledge of the significance of houses for their owners, the destructive effects of thawing snow, and the allusions to death create the immediate contextualisation.

The abundance of synecdochic detail in this passage conveys a degree of authorial choice in that the representation of the house is made through selection of visually prominent architectural features as parts for the whole. *Portico, eaves, parapet*, etc. all represent a huge mansion. A fictional scene is provided, enabling the reader to situate the context within the novel and, for Dickens' nineteenth century readership, with reference to a known world. A first reading may be thought to provide adequate interpretation. The strong implicature of the details in their literal sense of parts of a house seems to situate the scene, but further effort makes the reader aware of a wide range of weak implicatures created by the juxtaposition within the paragraph of literal with figurative language.

Several features move us away from the representational aspect and into the symbolic world created, from the start of this novel, through the move out of physical environment into interpretations of it. The selection of details is neither random nor all-encompassing. The single semantic field 'house' contains all of these elements and many more which the author ignores for his present purposes. Those which have been chosen and used as the focalising device are metonymic in character because through them the physical domain moves into the abstract. In the social code of this world, large houses with prominent architectural details stand for wealth, power and a sense of invulnerability. Then, the inferred metonym 'house for social status' is superimposed on a second complex metonym, the thawing snow. Literally bad weather which may destroy the house as it infiltrates the structure, the snow stands for forces in society, and events, which are figuratively attacking the representation of the house and what it stands for. Further, the snow is personified in the phrases *crept as if for shelter* and *wastes and dies*. The reader is reminded of his earlier fear that Lady Dedlock will meet this same fate (as indeed she does). This context is strengthened by the reference to Sir Leicester's other great house, Chesney Wold, with which the reader is already familiar. Its legend of a woman's death foretold by the sound of footsteps on the Ghost's Walk refers to Lady Dedlock.

The syntactic forms which Dickens uses include variants on normal prose discourse which evoke the poetic rhythms of an oral tradition of story-telling. The fronting of a series of prepositional phrases (*from the portico, into every chink... upon the skylight even through the skylight)* is enhanced by repetitions which have the effect of moving the synecdochic detail into a more abstract domain. Further poetic effects such as synaesthesia (a sensation of another kind suggested by one experienced) link the visual image to the relentless sound of the drips and by inference to the unpleasantness of thaw. Dickens wishes the reader to understand the symbolic importance attached to combining the house and the snow in a complex chain of metonyms which function with the syntactic form as a strong focalising device.

The passage is structured by using the principle of contiguity at more than one level. Firstly, connection with the rest of the novel is established through recalling the architectural details of the house, and the effects of London weather. Then, these spatio-temporal features are extended through the metonymy of cause-and-effect. A whole way of life, i.e. of the wealthy family represented by the mansion, is threatened. It works like this: 'house' stands for its owners, 'snow' stands for destruction, therefore the effects of snow on the house stand for the destruction of the owners. At another level, metonymic discourse is the trigger for an explosion of meaning as the reader is led away from the known world into abstract values and representations. 'Snow' has a multivalent implicature here; its function is to ground figurative elaboration in the reality of physical experience.

The question arises of whether there is any interest in adding a relevance-theoretic framework to the interpretations already available through rhetorical and linguistic analysis. From the relevance-theoretic point of view, the interpretation of the text must be subject to the principles of effort and effect. The reader's initial encounter with the text is based on the assumption that he will understand what the writer intends to communicate. The conventional synecdoches and perception of physical reality speed the process, acting as cues to what the reader is meant to understand. The first effect is to communicate clearly that snow is literally invading the house in a way which is common knowledge. The reader might stop here, as this is a plausible interpretation. There are clues in the text which encourage him to believe that the first interpretation is inadequate. But a further consideration is whether the use of metonymy here can simply be denied in favour of a literal reading, or explained as gratuitous, albeit interesting aesthetic effects. This interpretation might well be adequate insofar as the reader is satisfied that this is no more than an episode in the narrative, a moment of authorial comment which is easily accessible because the referents are all known to him already. The need to process the metonyms, by assigning referential value and enriching the scope of their linguistic form, reduces the relevance of the authorial communication by imposing gratuitous effort.

In response to this criticism, we may refer to Sperber and Wilson's (1986b) analysis of figurative language as 'Loose Talk'. Poetic effects such as metaphor, metonym and other figures of classical rhetoric are explained as a joint responsibility for speaker and hearer, or writer and reader. It seems possible in this example from Dickens to extend what the authors say about metaphor to metonymy:

> In general, the wider the range of potential implicatures and the greater the hearer's responsibility for constructing them, the more creative the metaphor. ... the hearer can go beyond exploring the immediate context and the

> background knowledge directly invoked, accessing a wider area of knowledge, entertaining ad hoc assumptions which may themselves be metaphorical, and getting more and more very weak implicatures.
>
> (Sperber and Wilson 1986b:168)

In the exemplification from *Bleak House*, the inferences made through the metonymic details are of great importance to understanding the scene and Rouncewell's premonitions of the death of Lady Dedlock, leading the reader to anticipation of the narrative. The potential implicatures reached through the metonymic discourse are, in relevance-theoretic terms, weakened but widened, offering more scope for communicating the author's intention. Relevance is created by shared properties: Rouncewell's observations as he walks around, the reader's knowledge of the linguistic structure of ghost stories, the content of the narrative, brought together through the metonyms.

3.2 Virginia Woolf: *To the Lighthouse*

The second example is an episode from a novel which appears to be structured in terms of metaphoric rather than metonymic discourse. Woolf's writing tends to place greater emphasis on metaphor and on aesthetic qualities of style than on forwarding the narrative. To find a clear example of metonymy is rare, but the episode of Mrs. Ramsay's necklace shows that Woolf, in this case, contextualises her highly figurative discourse and highlights its implied reality through using a striking metonym.

In this text, Mrs. Ramsay is dressing for dinner and allows her children Rose and Jasper to choose which necklace she is to wear. The narrator has shown that her mind is preoccupied with other things - the dinner itself, her desire to please the principal guest William Bankes, the late return from an outing of other children and guests, the antics of the rooks outside the window. Her thoughts are divided between her identity and duties as a hostess, her reflections on nature and people, and her role as mother. Spatio-temporal contiguity between Mrs. Ramsay, the children and the necklace in the first instance, and contiguity in the social code shared by author, narrator and reader provide the context which makes the referents accessible. This contextualisation leads the reader away from perceiving the necklace only as object, and towards other implicatures.

> But which was it to be? They had all the trays of her jewel-case open. The gold necklace, which was Italian, or the opal necklace, which Uncle James had brought her from India, or should she wear her amethysts?
>
> 'Choose, dearests, choose,' she said, hoping that they would make haste.
>
> But she let them take their time to choose: she let Rose, particularly, take up this and then that, and hold her jewels against the black dress, for this little ceremony of choosing jewels, which was gone through every night, was what Rose liked best, she knew. She had some hidden reason of her own for attaching great importance to this choosing what her mother was going to wear. What was the reason, Mrs. Ramsay wondered, standing still to let her clasp the necklace she had chosen, divining, through her own past, some deep, some buried, some quite speechless feeling that one had for one's mother at Rose's age.
>
> (V. Woolf, *To the Lighthouse* Ch.16, 307)

The narrative does not move forward until the choice has been made. The pause allows the reader to assign referential value to the object by various assumptions and inferences. The necklace has a number of strong implicatures. It is part of Mrs. Ramsay's outfit for the evening, thus standing for the self-image of the middle-class woman and associated with the social ritual of dressing for dinner. The reader is conscious of aesthetic pleasure in the objects (*'gold .... opal .... amethysts'*). The principal action of the scene, selecting the necklace, is encoded linguistically in the imperative *'Choose, dearests, choose,'* modified by two clauses *'hoping that'* .... and *'But she let them take their time'* .... The act of choice is controlled by Mrs. Ramsay, and stands metonymically for the power of the mother over the children.

The choosing of the necklace, foregrounded among the other elements of the narrative briefly, is a focalising device for one of the themes, i.e. Mrs. Ramsay's intuitive understanding of her daughter Rose with whom she has close links. The mother's acceptance of the choice forms a subtle metonym of cause and effect, standing for the transmission of mutual affection between mother and daughter. The use of Free Indirect Discourse gives access to Mrs. Ramsay's secret reflections and her explicit attribution of meaning: *'.... divining, through her own past, some deep, some buried, some quite speechless feeling that one had for one's mother at Rose's age.'*

The choice of Rose's preference rather than Jasper's encapsulates the mother's greater intuitive understanding of her daughter rather than her son. It represents the recurrent subtext in the novel that Mrs. Ramsay does not have as clear an understanding of men as of women. Rose is thought by her mother to attach great importance to selecting the necklace: *'.... was what Rose liked best, she knew.'* The necklace stands for an umbilical cord which neither mother nor daughter is ready to cut. The necklace, once chosen, becomes a metaphor, the link between the child and her mother. From the single domain of the mother's evening dress, it transfers into the two domains (jewellery and psychological dependence) of metaphor, but grounded in metonymy.

If the passage is examined in the framework of relevance theory, the verbal phrase *'Choose, dearests, choose'* satisfies both criteria for optimal relevance, i.e. for a minimum of processing effort there are maximum contextual effects. Within the fictional world, the children show that they have stopped at the first relevant interpretation, a literal one. The reader, however, influenced by the modifying clause *hoping they would make haste*, achieves a greater degree of relevant meaning. Thus, he finds clues in the adjacent text where the problem of reference to the characters' inner thoughts is disambiguated by explanatory comment: *'(Rose) had some hidden reason of her own for attaching great importance to this choosing what her mother was to wear.'*

The metonym, juxtaposing Mrs. Ramsay's overt wishes and hidden feelings, becomes the clue by which the reader is encouraged not to stop at the first interpretation. More processing effort is required than if the text simply read *Mrs. Ramsay allowed her children to choose a necklace and fasten it round her neck before dinner*. But a wider range of implicatures is accessed relevant to the narratorial point of view. These include the satisfaction felt when handling or wearing a beautiful object, the social importance of formal dress for dinner, the devotion of a girl to her mother and the mother's understanding and love for her, the memory of past experience, the dominant status of Mrs. Ramsay.

In this example, the need for an inferential interpretation of the metonym outweighs the value of a strictly rhetorical explication, and is in fact proposed by the author through the embedding of the metonymic object in text dominated by flashes of Mrs. Ramsay's 'stream of consciousness'. On the authorial side, the intention is clearly to signal that the text works at levels beyond the literal, for reasons to do with the coherence and the consistency of the narrative, the characterisation or the aesthetic concerns. Correct interpretation is enabled by metonymic representation bringing together a number of important concepts.

4. Conclusion

Investigation of metonymy as it occurs in everyday language indicates that a number of underlying principles distinguish this figure of speech from metaphor, but enable us to conflate it with synecdoche. In considering metonymy in 'literary' language, i.e. narrative prose fiction, similar principles are perceived. The concurrent presence of substitution, passage from the physical to the abstract, elliptical linguistic form, strong referentiality, contiguity, contextuality and inference have been demonstrated. The process of interpreting metonymy, or accounting for the gap between its linguistic form and semantic implications, is clarified by emphasising the importance of context and inference or, in relevance-theoretic terms, effect and effort.

The two examples from *Bleak House* and *To the Lighthouse* show that metonymy may be interpreted by means of analysing its linguistic form and its semantic implications in a given context. In this respect, the longer texts confirm the principles noted in examples of metonymy in everyday language. The first interpretation provided by the process of analysis and an understanding of the principles may be insufficient for full understanding. At this point a relevance-theoretic approach clarifies the process by which we can gain further insights, i.e. the pursuit of weak implicatures. Clues are given by both Dickens and Woolf as part of their communicative intention - if followed, they lead to a cornucopia of effects proper to figuratively enriched language. A successful approach to the meaning of these texts requires exegesis by rhetorical and linguistic means, and the use of the relevance-theoretic framework to give additional information about the process of interpretation.

References

Eco U. 1979. The Role of the Reader. London: Hutchinson.

Jakobson R. 1956. 'Two aspects of language and two types of aphasic disturbances'. In R. Jakobson and M. Halle. Fundamentals of Language. The Hague: Mouton 55-82.

Lakoff G. and M. Johnson. 1980. Metaphors We Live By. Chicago: University of Chicago Press.

Lodge D. 1977. The Modes of Modern Writing. London: Edward Arnold.

Nash W. 1989. Rhetoric. Oxford: Blackwell.

Preminger O. 1975. Princeton Encyclopaedia of Poetry and Poetics. London, Macmillan.

Schofer P. and D. Rice. 1977. 'Metaphor, metonymy and synecdoche revis(it)ed'. Semiotica 21, 1/2: 121-49.

Sperber D. and D. Wilson. 1986a. Relevance: Communication and Cognition. Oxford: Blackwell.

Sperber D. and D. Wilson. 1986b. 'Loose talk'. In Proceedings of the Aristotelian Society 1986: 153-71.

Wales K. 1989. A Dictionary of Stylistics. London: Longman.

Fictional texts cited:

Dickens C. 1853. Bleak House. Republished 1985. London: Penguin.

Woolf V. 1927. To the Lighthouse. Republished 1992. London: Penguin.

MEASURING SYNONYMY AS AN INTRA-LINGUISTIC AND CROSS-LINGUISTIC SENSE RELATION

Matutin Sikogukira (DAL)

Abstract

This paper discusses the sense relation of synonymy. It takes the view that the phenomenon of synonymy should be understood as a gradable concept, a 'cline' along which there are different degrees of synonymy. This view is consistent with the widely held opinion among semanticists that 'strict' or 'absolute' synonymy is rare in human language. A further step is taken to demonstrate that synonymy exists not only as an intra-linguistic sense relation but also as a cross-linguistic phenomenon. Thus, based on the criteria for synonymy, a more specific aim of this paper is to attest cross-linguistic synonymy, drawing evidence from French-English true cognates[1].

1. Introduction

Synonymy is one of the sense relations that semanticists have extensively written about. However, in spite of the amount of literature available on the phenomenon of synonymy, our understanding of it remains somewhat vague because it encompasses far more dimensions than our common sense actually perceives. This observation was made about two decades ago by Tutescu (1975:108) well before we saw the publications of the last one and a half decades or so in the field of semantics[2] :

> La synonymie est la relation sémantique qui a fait couler beaucoup d'encre, relation que le sens commun estime claire, mais que les logiciens ne cessent de proclamer crucifiante[3].

Synonymy exists as a phenomenon in everyday communication and in every language. Tze (1983) mentions two important functions that synonyms serve. First, they add flexibility to language by enabling its users to express the same meaning by different means. Second, they add variety and expressiveness to the language by enabling its users to exercise stylistic choices in conveying the same message. These two functions justify an investigation of synonymy as an intra-linguistic sense relation. Since sense relations normally hold between lexemes in the same language, looking at synonymy cross-linguistically is an unorthodox way adopted in this paper to try to illuminate the problem of classifying the so-called French-English true cognates. The question we need to answer in connection with the aim of this study is the following : Can French-English true cognates be described as synonyms, and to what extent? This discussion covers three stages. First, we define synonymy as a continuum with a view to demonstrating that 'strict' or 'absolute' synonymy is a rare

phenomenon in language. Second, we discuss the arguments put forward by semanticists against strict synonymy. Third, we examine a sample of French-English true cognates to attest cross-linguistic synonymy and to identify which criteria are useful for measuring synonymy.

2. Synonymy as a continuum : Defining criteria

We take as the starting point of our discussion two semantic intuitions. The first is that certain pairs or sets of lexical items bear a special sort of semantic resemblance to one another. It is customary to call items having this special similarity synonyms. The second intuition is that some pairs or sets of synonyms are more synonymous than others. In the introduction to *Webster's New Dictionary of Synonyms* (1968), it is made clear that, because there are too many factors involved in the selection of synonyms to make for absolute certainty or perfect accuracy in their choice, lexicographers do not always agree in their choice of synonyms. It is suggested that the only satisfactory test of synonymy is their agreement or likeness in *denotation*, even if this agreement is seldom so perfect as to make the words absolutely similar in meaning. There is, unfortunately, no neat way of characterising synonyms. It is obvious that synonyms must have a significant degree of semantic overlap, as evidenced by common semantic traits. However, it does not follow that the more semantic traits a pair of words share, the more synonymous they are. The assessment of synonymy rests on the nature of the differentiating characteristics because synonyms must not only manifest a high degree of semantic overlap, they must also have a low degree of implicit contrastiveness. In this respect, synonyms are lexical items whose senses are identical with regard to 'central' traits, but differ, if at all, only in respect of what we may describe as 'minor' or peripheral traits. This view seems to point to something like a scale of synonymy.

It is a widely held view that there are few, if any, 'real' synonyms in natural languages. To quote Ullmann (1957:108-9), 'it is a truism that total synonymy is an extremely rare occurrence, a luxury that language can ill afford'. Lyons (1981) proposes a scheme of classification which allows three possible kinds of synonymy :

(i) Full synonymy : synonyms are *fully synonymous* if, and only if, *all their meanings* are identical.

(ii) Total synonymy : synonyms are *totally synonymous* if, and only if, they are synonymous *in all contexts*.

(iii) Complete synonymy : synonyms are *completely synonymous* if, and only if, they are identical in *all (relevant) dimensions of meaning*.

According to Lyons's definitions, *absolute* synonyms are expressions that are fully, totally, and completely synonymous, whereas *partial* synonyms are synonymous but not absolutely so. Partial synonymy, which should not be confused with near-synonymy[4], meets the criterion of identity of meaning and is distinguished from absolute synonymy in terms of the failure of synonymous expressions to satisfy one or more of the conditions in (i), (ii), and (iii). He stresses that absolute synonymy, full synonymy, total synonymy, and complete synonymy (not to mention exact synonymy) are frequently employed as synonyms, whether absolute or partial, in

standard works, usually without definition. In general, it is complete and total synonymy that most semanticists have in mind when they talk of 'real' or 'absolute' synonymy but, in fact, it is true that there are very few such synonyms in language. The conclusion to draw from Lyons's discussion of the scale of synonymy is that some pairs or sets of synonyms are more synonymous than others, but, as will be shown in Section 4, there are difficulties to be contended with in the application of Lyons's criteria to pairs or sets of synonyms. In general, however, we tend to regard synonymy as a non-gradable concept.

3. Arguments against strict synonymy

Although the meanings of words may be the same or nearly so, there are three characteristics of words that rarely coincide : *frequency*, *distribution*, and *connotation*. Jackson (1988) presents two arguments against strict synonymy. One is economic, the other historical. Firstly, the economy of a language will not tolerate, except perhaps for a short period, the existence of two words with exactly the same range of contexts of use, and it certainly will not tolerate a proliferation of them. Secondly, historically, it has been noted that if strict synonyms occur in the language, whether by borrowing or for some other reason, then one of two phenomena tends to happen. The first phenomenon is that a differentiation of meaning takes place and one of the words begins to be used in contexts from which the other is excluded, perhaps through semantic specialisation. For example, Jackson (1988:66) points out that when *mouton* was borrowed into English from French, during the medieval period, it was absolutely synonymous with the Anglo-Saxon word *sheep*. Today, it still exists in the vocabulary of English as *mutton*, but with a specialised meaning referring to the meat of the animal consumed as food, while the animal is still called *sheep*. The second phenomenon is that one of the words in a synonym pair may be stylistically restricted. As far as borrowings into English from French are concerned, the borrowed word tends to be associated with more formal style. It is a well known fact that synonyms often differ in their etymological origin and stylistic use. Ullmann (1962:145-6) argues for this point in the following lines :

> ... There are in English countless pairs of synonyms where a native term is opposed to one borrowed from French, Latin, or Greek. In most cases the native word is more spontaneous, more informal and unpretentious, whereas the foreign one often has a learned, abstract, or even abstruse air. There may also be emotive differences : the 'Saxon' is apt to be warmer and homelier than its foreign counterpart ... There are many exceptions to this pattern; yet it recurs so persistently that it is fundamental to the structure of the language ...

Formality is but one dimension along which French-English true cognates differ as near-synonyms. And formality itself is far from being an all-or-none phenomenon but should rather be conceived as a 'cline' with various degrees of formality (see Tze 1983).

Finally, strict synonymy can also be countered by one of the words falling out of use or becoming obsolete or highly restricted, leaving the other as the sole lexeme with that meaning. For example, the word *enemy* was introduced into English from

Norman French, but English already had the word *foe* with the same meaning. Although *foe* is still employed in some contexts, mainly of a literary nature, *enemy* is much more used in most contexts. In British English, *foe* is regarded as an old-fashioned or formal word.

There is another dimension invoked by many semanticists in their discussion of synonymy, the distinction between 'cognitive' and 'emotive' or 'affective' synonymy. The former refers to the logical, cognitive or denotative content of a word and the latter refers to what is communicated of the feelings and attitudes of the speaker/writer. In the actual use of language, it is true that one word may be preferred to the other because of its emotive or evocative associations. However, the extent to which this is of importance varies considerably from one style or situation to another. For instance, the pairs *liberty/freedom* and *hide/conceal* are cited by Ullmann (1957) as examples of English words which are cognitively, but not emotively, synonymous. Although there are occasions when a speaker or writer might deliberately use one rather than the other of these synonyms, and make his choice on the basis of the 'connotations' that the words are likely to evoke, there are also many contexts in which either one or the other might be used without any noticeable difference of effect. Therefore we should not assume that the emotive connotations of a word are always relevant to its use. The truth is that in all cases, it is 'cognitive' synonymy which is defined first and no one ever talks of words as being 'emotively', but not 'cognitively', synonymous. Thus when we talk of synonymy, we do not generally have 'strict' or 'absolute' synonymy in mind. We are thinking much rather of pairs of words that can substitute for each other in a wide range of contexts but not necessarily absolutely, or that we think of as having the same general denotation or reference.

Since the description of meaning in a dictionary is an indication of the meaning potential of a word (only in a linguistic and situational context is the meaning actualised), synonymy also needs to be defined in terms of contexts of use. Two words are synonymous if they can be used interchangeably in all sentence contexts (see Jackson 1988; Lyons 1981 Ullmann, 1957). The pairs *discover/find, retain/keep*, and *occupied/busy* are commonly thought of as synonyms. Yet *discover* and *find* are synonymous in a sentence like 'We *discovered/found* the thieves hiding in a car park', but *find* could not substitute for *discover* in 'Sir Alexander Fleming *discovered/*found* penicillin in 1928'. Similarly, *retain* and *keep* are synonymous in the sentence '*Retain/keep* your ticket for further inspection', but *retain* could not substitute for *keep* in the sentence '*Keep/*retain* the door shut all the time'. The same observation applies to *occupied* and *busy* in the sentences 'The Prime Minister is *occupied/busy* at the moment' and 'The seat is *occupied/*busy*'. It appears therefore that 'likeness in denotation' is the most useful criterion for attesting synonymy.

4. Cross-linguistic synonymy between French-English true cognates

We need now to answer is the question we posed in the Introduction : Can French-English true cognates be described as synonym, and to what extent? To answer this twofold question, we need to show that certain French-English true cognates have a significant degree of semantic overlap, evidenced by common semantic traits, and

that, following the various degrees and types of synonymy discussed above, some pairs of French-English true cognates are more synonymous than others.

4.1 Problems of applicability of the criteria

The real problem lies in establishing some objective measure of the semantic overlap between French-English true cognate pairs. On the one hand, some of the definitions and types of synonymy discussed above make it impracticable to prove that two items are synonymous. First, Lyons's definition of 'total synonyms' as those which are synonymous in all contexts would require checking the relations between synonymous items in all conceivable contexts, which would be theoretically and practically impossible. Second, his definition of 'complete synonyms' as those which are 'identical on all relevant dimensions of meaning' leaves open the question of how many dimensions there are, and how to determine whether the words are identical on any particular dimension. Different analysts (semanticists) would not automatically agree on those two issues. Third, Lyons's definition of 'full synonyms' in terms of 'all their meanings being identical' poses a problem of knowing how many meanings each of the synonyms has. Fourth, it is not precisely clear where his category of 'near-synonyms', defined as those which are 'more or less similar, but not identical in meaning', would start and end because the phrase 'more or less' used in the definition is vague. Lastly, although he insists that near-synonymy is not the same as 'partial synonymy', he does not suggest a clear-cut criterion for differentiating between the two, and, by his definition, near-synonyms qualify as incomplete synonyms, and, therefore, as partial synonyms.

On the other hand, there are further difficulties to be contended with to attest synonymy and establish degrees of cross-linguistic synonymy. The first difficulty stems from the principles of the Sapir-Whorf Hypothesis, namely that language determines the way we think - *linguistic determinism* - and that the distinctions encoded in one language are not found in any other language - *linguistic relativity* - (cf. Mandelbaum 1949, Carroll 1956). Without going into the complex issue of what these principles imply, it is expected that the meanings of words in two languages rarely coincide totally, except highly technical words. Cognates are no exception to this rule. The second difficulty has to do with restrictions in the usage and distribution of cognates in two languages. Cognate lexemes may be more frequent, grammatically marked, sociolinguistically and collocationally restricted in one language than in the other.

4.2 Exemplification

With the above observations in mind, let us now examine an illustrative sample of French-English true cognates to see where they belong on the scale of synonymy in accordance with the criteria discussed Sections 2 and 3. The following examples will serve : *commencer/commence*, *restaurant/restaurant*, *inaugurer/inaugurate*, *succéder/succeed*, *terrible/terrible*, *brillant/brilliant*, *marchandise/merchandise*, and *sabot ige/sabotage*. We shall use the symbols N for normal, LN for less normal and AN for abnormal.

I. ***Commencer vs commence***

a. L'année académique/ *commence*/ en Octobre. (N)
b. The academic year/ *commences*/ in October. (N)

c. Le match/*commence*/ dans une heure de temps. (N)
d. The match/ *commences*/in an hour's time. (N)

e. Il *commence*/ à comprendre. (N)
f. He *commences*/ to understand. (LN)

g. Il *commence*/ à pleuvoir. (N)
h. It *commences*/ to rain. (LN)

In these pairs of sentences, the cognates *commencer* and *commence* are used synonymously but as we go down the list, we start getting a 'less normal' use of English *commence* whereas the use of French *commencer* is normal in all the four sentences. We have here an example of French-English cognates which share a common denotative and cognitive meaning but differ according to the register of formality, English *commence*, unlike French *commencer*, being mostly used in formal contexts. Therefore *commencer* and *commence* are not total synonyms because they are not synonymous in all contexts. They are partial synonyms. Additionally, there is a grammatical point to be made, namely that English *commence* rarely occurs with a non-finite clausal complement and that, when it does, the non-finite verb tends to be in the present participle form rather than the infinitive (e.g. They *commenced* eating). *Commence* seems to be mainly restricted to taking NPs as subject and object (e.g. The ceremony *commenced*/They *commenced* the ceremony).

II. ***Restaurant vs restaurant***

a. Nous allons/ manger/ dans un *restaurant*. (N)
b. We are going/ to eat/ in a *restaurant*. (N)

c. Ce *restaurant*/ est/ cher.
d. This *restaurant*/ is/ expensive.

e. Ils aiment/ les *restaurants* français. (N)
f. They like/ French *restaurants*. (N)

The cognate term *restaurant* has exactly the same meaning in all the three pairs of sentences and one cannot think of any context in which or any dimension on which they have different meanings. They satisfy the criteria for full synonymy, total synonymy, and complete synonymy and are therefore absolutely synonymous cognates.

III. ***Inaugurer vs inaugurate***

a. *Le Président Clinton/ a été *inauguré*/ le 20 Janvier 1993. (AN)
b. President Clinton/ was *inaugurated*/ on January 20th 1993. (N)

c. Le Premier Ministre/ *inaugurera*/ le Palais du Parlement. (N)

d. The Prime Minister/ will *inaugurate*/ the Parliament Building. (N)

In English you can inaugurate things as well as people but in French you can inaugurate things but not people. In French, when talking about 'the inauguration of people', the terms *investir* (to invest) and *investiture* (investiture) are used. Thus *inaugurer* and *inaugurate* fail the test of total synonymy because they are not synonymous in the two contexts. They are partially synonymous cognates.

IV. *Succéder vs succeed*

a. Qui/ va *succéder*/ à la reine? (N)
b. Who/ will *succeed*/ the queen?(N)

c. He/ has *succeeded*/ in his business. (N)
d. *Il/ a *succédé*/ dans ses affaires. (AN)

French *succéder* does not have the meaning that English *su ceed* has in (c), that is, 'to achieve the intended result or goal'. French uses a different lexeme *réussir* to express this meaning. Therefore *succéder* and *succeed* are not synonymous in all contexts and all their meanings are not identical. They are not totally or fully synonymous. They are not completely synonymous either because they are not identical on all relevant dimensions of meaning in that *succéder* in (d) does not have the denotative meaning that *succeed* has in (c). We cannot call them partial synonyms because they do not satisfy any of the criteria for full, total, and complete synonymy. We cannot call them near-synonyms because there is a meaning they do not share in (c) and (d). It is here that Lyons's distinction between partial synonymy and near-synonymy poses a problem. The cognates *succéder* and *succeed* share some but not all denotative meanings. They are synonyms of some kind. I propose to call them **incomplete synonyms** in contradistinction to Lyons's partial synonymy but incomplete synonymy is not the opposite of complete synonymy. Incomplete synonymy here refers to synonyms which differ by at least one denotative meaning. There are many other pairs of this type such as French *histoire* (which means both 'history' and 'story') and English *history* (which never means 'story') and French *siège* (which means 'siege, seat, and headquarters') and English *siege* (which does not have the last two meanings).

V. *Terrible vs terrible*

a. Quel/ *terrible* /accident! (N)
b. What/ a *terrible*/ accident! (N)

c. Tous les Marseillais/ regardaient/ la Télévision/ quand/ leur équipe/ a l'emporté/ la Coupe d'Europe. Le match/ était/ *terrible*. (N)
d. All people from Marseilles/ were watching/ Television/ when their team/ won/ the European Cup. The match/ was/ *terrible*. (AN)

French *terrible* in (c) has an emotive meaning of 'great' which English *terrible* does not have in (d). This meaning is expressed by *terrific* in English. English *terrible* also has an emotive meaning which can be the opposite to the French meaning as in 'Le film était *terrible*' (the film was terrific) and 'The film was *terrible*'. In these

examples, the item *terrible* can be viewed as an isolated example of **cross-linguistic antonymy**.

VI. *Brillant vs brilliant*

a. C'est/ un étudiant *brillant*. (N)
b. He is/ a *brilliant* student. (N)

c. Elle/ a/ une carrière *brillante*. (N)
d. She/ has/ a *brilliant* career. (N)

e. *Comment/ était/ votre congé? C'était/ *brillant*. (AN)
f. How/ was/ your holiday? It was/ *brilliant*. (N)

In (f) English *brilliant* has an emotive meaning of 'fantastic' which French *brillant* does not have in (e). I propose to call the pairs *terrible/terrible* and *brillant/brilliant* incomplete synonyms because they share their cognitive meaning but not their emotive meaning.

VII. *Marchandise vs merchandise* [5a]

a. Cette *marchandise*/ est chère. (N)
b. This *merchandise*/ is expensive.(N)

c. Ces *marchandises*/ sont/ chères.(N)
d. *These *merchandises* are expensive.(AN)

The use of *merchandise* in (d) is abnormal because *merchandise* is an uncountable noun, unlike French *marchandise*, which is a countable noun. We have here a pair of French-English true cognates which differ in their grammatical meaning. They are partial synonyms because although their denotative meaning is identical (we can therefore call them full synonyms), they are not synonymous in all contexts because of their 'count/uncount differentiation.

VIII. *Sabotage vs sabotage* [5b]

a. Ils/ veulent/*saboter* (**sabotage*)/ les Jeux Olympiques. (N)
b. They/ want/ to *sabotage*/ the Olympic Games. (N)

c. Le *sabotage* des négociations/ va/ continuer. (N)
d. The *sabotage* of the negotiations/ will/ continue. (N)

Owing to the phenomenon of conversion[6], English *sabotage* functions as both a verb and a noun whereas French *sabotage* is a deverbal noun from the verb *saboter* (to sabotage). Therefore French *sabotage* and English *sabotage* are not synonymous in all contexts and so are not total synonyms; they are not full synonyms because all their meanings are not identical and they are not complete synonyms because they are not identical on the dimension of their grammatical meaning (synonyms must belong to the same word class). They are not partial synonyms. They are incomplete synonyms because of their word class differentiation.

5. Conclusion

The above discussion has demonstrated that it is erroneous to think of synonymy as a monolithic phenomenon. Synonymy covers so many dimensions and aspects of semantic equivalence that its measurement is more complex than it appears to be. It has also been shown that synonymy is an important cross-linguistic sense relation between French-English true cognates. However, as is true of intra-linguistic synonyms, many French-English true cognates are 'partial' or 'incomplete' synonyms and only a few of them qualify as 'strict' or 'absolute' synonyms. From the sample of cognates discussed above, it can be concluded that 'agreement' or 'likeness' in denotation is the most useful criterion for measuring cross-linguistic synonymy. Most French-English true cognates share their denotative meaning but tend to differ in terms of formality, emotive meaning, and grammatical traits and it is these dimensions that run counter to the criterion of 'absolute semantic equivalence' of cross-linguistic synonyms.

Notes

1. In the area of lexis, cognates are items which exist in two or more languages, always present some resemblance in form (orthographically, phonologically and/or phonologically, and morphologically), are usually but not always etymologically related, and may but need have similarity in meaning. There are two main categories of cognates, commonly known as true cognates and false cognates. True cognates are words which are etymologically related and whose meanings and ranges of meanings completely or almost completely overlap, e.g. English *hotel* and Spanish *hotel*, English *restaurant* and French *restaurant*. False cognates are words which are etymologically related but whose meanings and ranges of meaning do not overlap, e.g. English *tutor* (lecturer or teacher) and French *tuteur* (guardian), English *auditorium* (place for gathering) and Spanish *auditorio* (an audience).

2. See the following collections: Cruse 1986; Hurford and Heasley 1983; Lyons 1977; Palmer 1976/1981).

3. Synonymy is the semantic relation that has so much been written about, a relation that common sense perceives as clear but which logicians constantly find 'crucifying'.

4. Near-synonyms are 'expressions that are more or less similar, but not identical in meaning' (Lyons 1981:50).

5a and 5b. The items *marchandises/merchandise* and *sabotage/sabotage* belong to a separate category of cross-linguistic synonyms from the preceding types because the English and French forms differ only in grammar. These items do not collocate syntagmatically with the same range of other lexemes and so are not synonymous in all contexts. On the one hand, French *marchandise* is a countable noun but English *merchandise* is not. On the other hand, the pair *sabotage/sabotage* is only synonymous insofar as they belong to the same word class. In English, the noun *sabotage* and the verb *sabotage* are the same orthographic word but are separate lexemes because they belong to different word classes.

6. *Conversion* is the change in word class of a word without any corresponding change in form. That is, a stem is derived without any change in form from one belonging to a different class (see Bauer 1983; Huddleston 1988).

References

Bauer L. 1983. English Word-Formation. Cambridge : Cambridge University Press.

Carroll J. (ed.). 1956. Language, Thought and Reality : Selected Readings on Benjamin Lee Whorf. New York : John Wiley.

Cruse D. 1986. Lexical Semantics. Cambridge : Cambridge University Press.

Huddleston R. 1988. English Grammar : an Outline. Cambridge : Cambridge University Press.

Hurford J. and B. Heasley. 1983. Semantics : A Coursebook. Cambridge : Cambridge University Press.

Jackson H. 1988. Words and their Meaning. London : Longman.

Lyons J. 1981. Semantics. vol.1 & 2. Cambridge : Cambridge University Press.

Lyons J. 1981. Language, Meaning and Context. London : Fontana.

Mandelbaum D. (ed.). 1949. Selected Writings on Language, Culture and Personality. Berkeley : University of California Press.

Palmer F. 1976/1981. Semantics. Cambridge : Cambridge University Press.

Tutescu M. 1975. Précis de Sémantique Française. Paris : Klincksieck.

Tze C. 1983. 'Towards an Explication and Description of Synonymy in English'. Ph.D Thesis. University of Edinburgh.

Ullmann S. 1957. The Principles of Semantics. Oxford : Basil Blackwell.

Ullmann S. 1962. Semantics : An Introduction to the Science of Meaning. Oxford : Basil Blackwell.

Webster's New Dictionary of Synonyms. 1968. Massachusetts : G. & C. Merriam Co.

ial

Issues in Applied Linguistics

Issues in Applied Linguistics (IAL) is a refereed scholarly journal presenting research in the areas of language acquisition, discourse analysis, language education, neurobiology of language, language testing, language and culture, and research methodology. *IAL* particularly welcomes submissions from graduate students.

Upcoming issues of *IAL*:

Volume 5.1: June 1994

(Analysis of Classroom Discourse, Corpus Based Grammatical Analysis).

Volume 5.2: December 1994

(Applied Linguistics from an East Asian Perspective).

Subscription Information

The one year subscription rates (two issues) are as follows:

	USA/Canada	Other Countries*
Students	US$15.00	US$20.00
Faculty	US$25.00	US$30.00
Institutions	US$35.00	US$40.00

*For subscriptions outside of the US or Canada, the additional fee is for overseas mailing costs (Surface Mail). Add $9.00 for Air Mail.

Make check or money order payable (in U.S. funds drawn on a U.S. bank, encoded for U.S. Federal Reserve System) to *Issues in Applied Linguistics*.

IAL is abstracted in *Linguistics and Language Behavior Abstracts (LLBA)*, is indexed on *Comserve*, and is listed in the *MLA International Bibliography*.

Back issues are available for purchase through IAL.

Submissions/Subscriptions/Advertisements:

Issues in Applied Linguistics
UCLA Department of TESL & Applied Linguistics
3300 Rolfe Hall
405 Hilgard Avenue
Los Angeles, CA 90024-1531
USA

Forget About Needles in Haystacks!

Discover the information you need quickly and easily by consulting **Linguistics and Language Behavior Abstracts (LLBA)**—your window on the entire spectrum of linguistic and language-related research.

As a specialist in your field, it is important to keep pace with the latest research findings. **LLBA** can help you do just that with timely coverage of articles culled from among 1,900 serials, selected books, and enhanced bibliographic citations for relevant book reviews and dissertations. **LLBA** offers high-quality abstracts, precise indexing, and comprehensive backfiles.

Among the subject areas covered are:

- Applied and Theoretical Linguistics
- Descriptive Linguistics
- Interpersonal Behavior and Communication
- Psycholinguistics
- Nonverbal Communication and more

LLBA *is available in formats suitable to every setting:*

- in print (5 times a year)
- online from BRS, DIALOG, and now, on CD-ROM—**LLBA Disc**—with SilverPlatter's retrieval software.

Our support services include:

- The *Thesaurus of Linguistic Indexing Terms*
- The *LLBA User Reference Manual*

for further information about our products and services contact:

Linguistics and Language Behavior Abstracts

P.O. Box 22206, San Diego, CA 92192-0206

(619) 695-8803 FAX (619) 695-0416 Internet: socio@cerf.net

Institute for Applied Language Studies (IALS)

The Institute's English courses are designed to meet specialist language learning needs and run all year round. Specialist courses are available for law, business and management, medicine and also academic preparation.

As well as English courses, IALS is a centre for the teaching of other languages; R & D; distance learning; consultancy and courses abroad; publications and materials for language learning; worldwide language development projects such as the Edinburgh Project in Extensive Reading.

IALS

a creative language learning environment

Institute for Applied Language Studies
21 Hill Place, Edinburgh, EH8 9DP Tel: 031-650 6200
Telex: 727442 UNIVED G Fax: 031-667 5927

TEACHERS' COURSES AT IALS

The Institute for Applied Language Studies offers a wide range of courses and other services for teachers of English and other languages.

Certificated courses (full-time)

- UCLES/RSA Diploma in the Teaching of English as a Foreign Language to Adults (one term, commencing October and March)
- Advanced Certificate in English Language Teaching (one term, commencing October and March)
- Advanced Certificate in English Studies (two terms, commencing October)
- Higher Advanced Certificate in English Studies (three terms, commencing October)

Certificated courses (distance-learning)

- Advanced Certificate in English Language Teaching (one year, commencing June for individual enrolments)
- RSA Diploma in the Teaching of Foreign Languages (one year, commencing September)

Short summer courses

- Teaching and Learning English
- Teaching English for Specific Purposes
- Teaching English for Medical Purposes
- Teaching English for Business Purposes
- Grammar and the Communicative Teaching of English
- Drama for TEFL
- CEELT Preparation course
- English for EFL/Applied Linguistics

Trainer Training

- A modular course, available only to sponsored groups

Tailor-made versions of these courses, together with seminars on specialised topics and a language-teaching consultancy service, can also be supplied overseas.

For further information write to "Teachers' Courses",
IALS, University of Edinburgh, 21 Hill Place, Edinburgh EH8 9DP.
Tel. 031 650 6200 Fax. 031 667 5927

Contents of previous issues: EWPAL 1 (1990)

FALLING FROM GRICE: THE IDEOLOGICAL TRAP OF PRAGMATIC STYLISTICS
Peter Tan

UG, SLA AND LANGUAGE TEACHING: FUTURE OF AN ILLUSION?
Liam Rodger

WHAT, IF ANYTHING, IS ENGLISH FOR LITERARY STUDIES?
Brian Parkinson

TRAINER TRAINING, SIMULATION AND THE FEEDBACK IMPERATIVE
Ian McGrath and Sevgi Altay

THE WASHBACK EFFECT OF A TEXTBOOK-BASED TEST
T R Khaniya

TALKS: WHAT NEGOTIATORS THINK DRIVES A HARD BARGAIN
Mary Ann Julian

THE EFFECT OF CO-OPERATIVE TEACHING ON THE QUANTITY AND QUALITY OF TEACHER FOREIGN LANGAUGE DISCOURSE IN THE FRENCH CLASSROOM
Carole E M Franklin

DO THE SIMIALRITIES BETWEEN L1 AND L2 WRITING PROCESSES CONCEAL IMPORTANT DIFFERENCES?
Ana Frankenberg-Garcia

COURSE DEVELOPMENT IN NATURALISTIC PERSPECTIVE
Gibson Ferguson and Gillian D Brown

HAVE THE GOAL POSTS BEEN MOVED? A consideration of goals of African speakers of English in the context of the emergence of non-native varieties of English
Esther Daborn

SER ou NAO SER? A study of Spanish-Portuguese cross linguistic influence
Cathy Benson

Contents of previous issues: EWPAL 2 (1991)

DISCURSIVE ASPECTS OF METAFICTION: A NEO-ORAL AURA?
Alexandra Georgakopoulou

THE REACTION OF LEARNERS TO TAPE-BASED LISTENING COMPREHENSION MATERIALS IN FRENCH, SPANISH AND ITALIAN 'COMMUNITY CLASSES'
Brian Parkinson, Giulia Dawson, Lucila Makin, Hélène Mulphin

INTERLANGUAGE LEXIS: AN INVESTIGATION OF VERB CHOICE
David J. Hill

AN INVESTIGATION OF A TIMETABLED SELF-ACCESS SESSION IN A GENERAL ENGLISH PROGRAMME
Sheena Davies, Eileen Dwyer, Anne Heller, Kate Lawrence

IS IT OR IS IT NOT INTERLANGUAGE? A HEAD-ON CONFRONTATION WITH NON-NATIVE ENGLISH
L.K. Owusu-Ansah

CONDITIONALS AND THE EXPRESSION OF REGRET AND RELIEF: TOWARDS A FRAGMENT FOR A COMMUNICATIVE GRAMMAR
Gibson Ferguson

ARGUMENTS FOR AND AGAINST FREE VARIATION
S.B. Makoni

BIBLIOGRAPHIC PRESENTATION
Tony Lynch and Ian McGrath

CROSSLINGUISTIC INFLUENCE IN A BILINGUAL CLASSROOM: THE EXAMPLE OF MALTESE AND ENGLISH
Antoinette Camilleri

ASSESSING THE READABILITY OF MEDICAL JOURNAL ARTICLES: AN ANALYSIS OF TEACHER JUDGEMENTS
Gibson Ferguson and Joan Maclean

LITERARY DISCOURSE AND IRONY: SECRET COMMUNION AND THE PACT OF RECIPROCITY
Sonia S'hiri

Contents of previous issues: EWPAL 3 (1992)

ASYMMETRIC RESETTING OF THE NON-EMPTY TOPIC PARAMETER BY CHINESE-SPEAKING LEARNERS OF ENGLISH

Boping Yuan

THE SOCIOLINGUISTIC STATUS OF ENGLISH IN MALTA

Antoinette Camilleri

INTERLANGUAGE PHONOLOGY: THE PERCEPTUAL DEVELOPMENT OF DURATIONAL CONTRASTS BY ENGLISH-SPEAKING LEARNERS OF JAPANESE

Kayoko Enomoto

PARTICIPANT REFERENCE IN GREEK PERSONAL EXPERIENCE NARRATIVES

Alexandra Georgakopoulou

READING, CULTURE AND COGNITION

Martin Gill

CLUSTER ANALYSIS AND THE INTERLANGUAGE LEXICON

David J. Hill

INTERACTIONAL LISTENING TASKS: A COMPARATIVE STUDY OF STRATEGY AND PRACTICE TEACHING APPROACHES

Vanessa Luk

LESSON BEGINNINGS

Ian McGrath, Sheena Davies and Helène Mulphin

WHERE TRUE POWER LIES: MODALITY AS AN INDICATION OF POWER IN TWO INSTITUTIONALISED DOMAINS OF LANGUAGE USE

L.K. Owusu-Ansah

TENSE, ASPECT AND THE BUSCONDUCTOR HINES - THE LITERARY FUNCTION OF NON-STANDARD LANGUAGE IN THE FICTION OF JAMES KELMAN

Liam Rodger

A PRAGMATICS OF VERBAL IRONY IN LITERARY DISCOURSE: AN EXAMPLE FROM DRAMA

Sonia Shiri

AN INVESTIGATION INTO LEARNERS' REACTIONS TO LEARNER TRAINING FOR SELF-ACCESS LISTENING

Elspeth Wardrop and Kenneth Anderson

Contents of previous issues: EWPAL 4 (1993)

MSC COMMON ROOM CONVERSATIONS: TOPICS AND TERMS
Joan Cutting

SPECULATION AND EMPIRICISM IN APPLIED LINGUISTICS
Alan Davies

SEASPEAK
Alan Davies

IMPLEMENTING INNOVATION IN LANGUAGE EDUCATION
Gibson Ferguson

FUNCTIONAL CONTROLLED WRITING
Ardeshir Geranpayeh

DEVELOPMENT AND VALIDATION OF A TRANSLATION TEST
Behzad Ghonsooly

THE SIGNIFICANCE OF 'SIGNIFICANCE'
Martin Gill

THE DESIGN AND VALIDATION OF A MULTI-LEVEL READING COMPREHENSION TEST
Aileen Irvine

QUESTIONS IN LECTURES: OPPORTUNITIES OR OBSTACLES?
Tony Lynch

CAN APPLIED LINGUISTS DO ETHNOGRAPHIC INTERVIEWS?
Brian Parkinson

INFLUENCE OF LANGUAGES OTHER THAN THE L1 ON A FOREIGN LANGUAGE: A CASE OF TRANSER FROM L2 TO L3
Matutin Sikogukira

EWPAL 5 CONTRIBUTORS' ADDRESSES:

Joan Cutting Alan Davies Ardeshir Geranpayeh Anne Pankhurst Carmen Santos Maldonado Matutin Sikogukira	*Dept. of Applied Linguistics* *14 Buccleuch Place* *Edinburgh EH8 9LN* *Scotland - UK*

Sheena Davies Tony Lynch Brian Parkinson	*IALS* *21 Hill Place* *Edinburgh EH8 9DP* *Scotland - UK*

Kayoko Enomoto	*Centre for Asian Studies* *University of Adelaide* *South Australia 5005* *Australia*

SUBSCRIPTION / EXCHANGE FORM FOR EWPAL

Name/Institution

Address

Please tick the relevant items:

Issue number *	EWPAL number 2 (1991)		.....
	EWPAL number 3 (1992)		.....
	EWPAL number 4 (1993)		.....
	EWPAL number 5 (1994)		.....
	EWPAL number 6 (1995)		.....
Type of subscription (price per issue)	individual: surface / EC	£5.50	.....
	individual: non-EC air mail	£7.50	.....
	institutional: surface / EC	£6.50	.....
	institutional: non-EC air mail	£8.50	.....
	institutional exchange		.....

(For exchanges) I enclose a current copy of our institutional publication entitled .....

.....

Total amount of payment: £ .....

Method of payment

Sterling cheque (enclosed) to University of Edinburgh .....

Sterling transfer (reference no. .....)
to the Royal Bank of Scotland, Nicolson Street, Edinburgh,
Scotland - bank code 83-19-19, account IALS 00181202 .....

Please send this form completed to : EWPAL subscriptions
IALS
21 Hill Place
Edinburgh EH8 9DP
Scotland - UK

* EWPAL number 1 (1990) is now out of print, but copies of single articles are available on request, at an additional charge (£2.50), to subscribers to any of the subsequent issues.

ISSN 0959-2253